

The international journal of science / 7 May 2020

# nature



## A TRICK OF THE TAIL

Fossil remains reveal aquatic propulsion and predation in a dinosaur

### COVID-19

A profile of the pandemic's killer coronavirus

### Exotic atom

Generation and spectroscopic analysis of pionic helium

### Alzheimer's disease

Genetic risk factor linked to disruption of blood-brain barrier

Vol. 581, No. 7806  
nature.com







## EPA leadership must be held to account

**The agency continues to aggressively undermine environmental regulations and demoralize its staff.**

**W**hy are those at the helm of the US Environmental Protection Agency (EPA) continuing to roll back key environmental and public-health regulations? Why are they making it harder for themselves – and future administrations – to craft new regulations based on all of the available research? And why, in their actions, do they continue to challenge the place of rigorous evidence in policymaking?

We need to keep asking these questions, because, although the world's attention is rightly focused on the coronavirus pandemic, in the past six weeks alone, the agency's leaders have accelerated moves to weaken research-based regulations.

They are walking a road that the EPA's founders – who launched the agency 50 years ago – could not have imagined the organization would take. In spite of reasoned opposition from researchers and many businesses, too, the EPA's leadership – with the tacit support of many Republicans in the US Congress – is undermining the very reasons for its own existence. This is a travesty.

The latest target is the well-established body of evidence that demonstrates how tiny airborne particles embed themselves deep in the lungs and increase the risk of premature death from respiratory and cardiovascular diseases.

In a report published last September, the EPA's staff reviewed recent literature, including research indicating that even low levels of particulate pollution contribute to premature deaths in the United States. The report concluded that such evidence justified reducing the maximum permitted concentrations of fine particulate matter from 12 micrograms per cubic metre of air to between 8 and 10 micrograms. But on 14 April, the agency's administrator, Andrew Wheeler, said that there is no need for change. No reasoning was provided for this decision except that: "The U.S. has made incredible strides in reducing particulate matter concentrations across the nation."

Two days later – in a different sphere of regulation – the EPA confirmed a change to the valuation methodology for regulations curbing the emissions of mercury and other pollutants from fossil-fuel-fired power plants – a change that could have far-reaching consequences for people and the environment.

Enforcing existing mercury standards costs the US

utilities industry up to US\$10 billion annually. In 2015, the administration of Barack Obama calculated that, in addition to preventing up to 11,000 premature deaths in the United States, these standards also save the country in the range of \$37 billion to \$90 billion a year – money that would otherwise be spent on, for example, extra health-care costs. But the EPA's leadership brushed aside the research and data that support these findings. The EPA is currently considering new guidance that would allow it to ignore such comprehensive assessments of comparisons of costs versus benefits – leading to potentially weaker environmental and public-health regulations in other areas, too.

On top of all this, in March the administration of US President Donald Trump finalized a revised set of targets governing greenhouse-gas emissions from new vehicles. It slashed the original ambition of 5% in annual reductions to 1.5%. (By contrast, the European Union's target is for a 27% reduction, compared with its 2015 target, in carbon emissions per kilometre from cars manufactured after 2020.)

### Controlling the science

Each of these revisions builds on the Trump administration's overarching narrative since 2017: to control the research that is used as evidence for the EPA's decisions. This began with a proposed rule that is misleadingly called Strengthening Transparency in Regulatory Science. In reality, it is a declaration that the EPA intends to limit its use of studies for which underlying data and models are not publicly accessible. This move has provoked widespread condemnation from researchers and searching questions from the EPA's own science advisory board. The rule, if implemented as written, could allow the agency to disregard landmark public-health studies for which underlying data – such as confidential health records – are not accessible for good reasons.

In all, more than 80 EPA rules and regulations are being revised or removed, and researchers have been campaigning and protesting as never before. They must continue to do so, working with allies in industry and across other sectors. It can be difficult and frustrating when the doors to reasoned engagement are closing. But it is, nevertheless, crucial to push back, so that the agency can be rebuilt when – not if – the time comes.

When the EPA disbanded one of its air-quality science advisory panels in 2018, the panel's members reconvened independently to continue their work, and issued a statement advising the agency to strengthen the particulate-matter standard. And when the agency refused to hold an evidence session on its proposed 'transparency' rule, members of the Union of Concerned Scientists, a science advocacy group based in Cambridge, Massachusetts, organized their own public hearing.

Some familiar names in industry also no longer wish to be associated with regulatory decisions they know will be harmful – and are choosing to stand with the science. Automobile manufacturers Ford, BMW, Honda and Volkswagen have said that they wish to comply with the stricter vehicle-emissions standards developed by the state of California. Some power corporations that have

**In all, more than 80 EPA rules and regulations are being revised or removed."**



already invested in meeting existing regulatory standards are also unwilling to roll back.

The EPA is becoming a shell of its former self. Its leaders have chosen to abdicate leadership, disregard evidence and expose the country's environment and health to risk of further degradation. Many staff have understandably chosen to leave. Those who remain should know that they have supporters around the world, rooting for them and for the EPA – and hoping that, one day, it will once again become an evidence-based body that lives up to its name and secures its founders' legacy.

## Share lessons on lifting lockdowns

**Countries are starting to lift restrictions on movement, but there is little consensus on how this should be done.**

**W**hen can lockdowns be lifted, and how should this happen? These urgent questions are being asked in each of the 180 or so countries and territories that enacted full or partial restrictions on movement in response to the new coronavirus. But as countries start tentatively to reopen schools, businesses and public places, it is also becoming clear that there is little consensus on how this should be done, because efforts to coordinate actions globally are running into the ground.

The need to get countries back to work is, of course, urgent. According to the International Labour Organization (ILO), an agency of the United Nations, stay-at-home orders have imperilled the lives and livelihoods of at least 1.6 billion people. The majority are workers who have no income or health protections and whose work cannot be done from home. No work means no food, and no money to pay rent. Governments have stepped in with various forms of temporary financial assistance, but they have also taken on large debts to provide this support.

Researchers are warning that when this financial support ends, the ill health induced by poverty could worsen. The World Food Programme is warning that 265 million people will experience acute hunger by the end of 2020 – twice the number facing food crises before COVID-19 (see [go.nature.com/3c8aebj](https://go.nature.com/3c8aebj)).

All of this is adding to pressure to reopen economies and societies. Yet the pandemic has not yet run its course, and infections and deaths continue to rise. If restrictions are lifted prematurely, the impacts of the virus risk being prolonged.

The World Health Organization published guidance last month for countries considering easing some restrictions

– advice that is intended to minimize a risk of a resurgence of the virus (see [go.nature.com/2a4jdmr](https://go.nature.com/2a4jdmr)). There can be no one-size-fits-all solution to easing restrictions. However, guided by systematic reviews of the peer-reviewed literature, and by its decades of experience in dealing with infectious diseases, the agency recommends that, in any scenario, disease transmission must be controlled – otherwise, without a vaccine, there's a risk that the virus will return. In many countries, transmission is uncontrolled and must be reduced so that cases arise only sporadically, or in clusters in identifiable regions. This, in turn, requires public-health systems to detect, test, isolate and treat cases of infection, and to trace people who had been in contact with those infected – something that would require large teams of contact-tracers.

The WHO's guidance, in addition, says that countries need to minimize transmission in potential hotspots, such as care homes or large gatherings; and that places where people gather – for example, schools and workplaces – should be provided with equipment to ensure that physical distancing can be implemented. The agency is also clear – as is the ILO – that governments must continue to provide financial and other forms of support to protect the most vulnerable groups.

By 1 May, according to researchers at the University of Oxford's Blavatnik School of Government, UK, no country had met even four of the six WHO criteria that a 150-member team convened by the university has been measuring. And only 20 countries and territories, including Trinidad and Tobago, Croatia, Hong Kong and South Korea have come close (see [go.nature.com/2zbs1ef](https://go.nature.com/2zbs1ef)).

This is largely because most countries (more than 150) have not yet met the first requirement – for COVID-19 to be reduced to sporadic cases and identifiable clusters. Moreover, at least 50 countries still lack adequate policies for testing, contact-tracing and isolating infected individuals. Such measures, along with physical distancing, have also been essential in helping to keep China's cases under control since travel restrictions were lifted in February – as research published this week in *Nature* confirms (S. Lai *et al.* *Nature* <https://doi.org/10.1038/s41586-020-2293-x>; 2020).

On 4 May, the international community – led by European Union member states and global philanthropic funders – raised US\$8 billion for research and development on COVID-19 and pledged to work with the WHO to accelerate efforts to develop vaccines and drugs. Such collaboration is essential and is to be welcomed, because it will ensure that communities in all countries can eventually be protected from the disease. But paying attention to the WHO's guidelines, and sharing best practice on easing lockdowns, is essential, too.

Countries are having to make difficult decisions as lockdowns are eased. And they are doing so while their researchers are still gathering and processing data. These data should be published and shared. Circumstances in individual countries will necessarily be different, but the world will benefit from the mutual learning that will come from such a global effort.

**“Collaboration will ensure that communities in all countries can eventually be protected from the disease.”**



# World view

## Let COVID-19 expand awareness of disability tech



By Ashley Shew

**The pandemic's disruption shows how much academia could learn from the disability community.**

**D**isabled people including myself have long campaigned for accommodations to help us live our lives. The COVID-19 pandemic shows that these are not as impractical as we have always been told. Supermarkets, restaurants and pharmacies (even outside cities) can deliver; remote working, medicine and education are possible for many; and social lives can be rewarding without requiring us to leave home.

All around me, I see academic colleagues adopting disability-led hacks and long-sought accommodations. I wish everyone had thought about these workarounds – and about disabled people at all – earlier. When lockdowns end, we must not forget these lessons, not least because the pandemic will disable people, and the impacts will be felt most by the most vulnerable parts of society.

Academia is paying for its ableism. At many universities, in-person research with human participants and in laboratories has been curtailed. If these projects had considered disabilities, they might be better off: disabled academics already plan in short increments, with built-in flexibility.

In 2014, I returned to my job as an assistant professor, newly multiply-disabled – a hard-of-hearing amputee battered by chemotherapy and more. I felt out of place. I could no longer access many spaces, including most of my colleagues' offices, and I sought the camaraderie of other disabled faculty members, staff and students. My disabled comrades and I recognize the diversity of disabilities: congenital and acquired; ranging from cognitive to sensory, mobility and more; apparent and not. Many of us pursue research that emphasizes how disabled people are the best experts on the technologies and structures that meet our needs.

One of my projects examines accounts of disabled people's lived experiences with technologies, and how they differ from those of the scientists and engineers behind the tech. I planned my work intending to recruit disabled students among my researchers. Most of the studies can be done remotely, even from bed, and on a funky, asynchronous timetable as needed. Last year, when I had lung surgery, my group shifted gears without worry. And because of its disability-led design, my team's project is pandemic-proof.

Another project, to gauge the experiences of students in civil-engineering classes, was designed to include participants with a range of disabilities. So we obtained approval for flexibility in communication format: we conduct our interviews by text, e-mail, Zoom and other means. Because we planned for disabled people to lead and participate in

  
**Academia is paying for its ableism."**

**Ashley Shew** is an assistant professor in the Department of Science, Technology, and Society at Virginia Polytechnic Institute and State University in Blacksburg. e-mail: shew@vt.edu

the research, we're well prepared for the current situation – or for any other.

Many disabled people are also adept at managing our energy, and forgiving ourselves for not always meeting conventional metrics of 'productivity'. My non-disabled colleagues are now struggling to adjust, but my team appreciates that 'clocks should bend to our bodies', not the other way around. Some disabled people call this concept Crip Time, reclaiming a derogatory term in pride (much like 'queer' for many LGBT+ people).

The disability community creates and lobbies for technologies and infrastructure that work better for all. Deaf and disabled people fought hard for things such as captioning on television, which has since become ubiquitous in sports bars and airports and can now be appreciated by people streaming media while those they live with rest or work.

The bitter irony is that, at the moment when non-disabled – or not-yet-disabled – people are beginning to normalize these disability hacks and hard-won infrastructure, society's disregard for disabled people is clear. We are dying of COVID-19 in greater numbers than are non-disabled people, in rehabilitation facilities, state institutions (including prisons), group homes and care homes.

Many accommodations demanded under COVID-19 were implemented within weeks, including the ability to work from home, to have flexible schedules, to get what we need without excessive and demeaning documentation, to share and celebrate creative adaptation, to work with the knowledge that all schedules can change. These are all things that disabled and chronically ill people have wanted for a very long time. I hope that when we've flattened the curve and saved as many people as possible, we don't return to a world in which disabled people are ignored (especially when COVID-19 will probably produce more of us).

So start making changes that should have been standard all along. Plan creatively and accessibly to allow more work offsite, and to include people whose clocks aren't steady. Welcome suggestions from disabled colleagues and students about how to make the environment work best for their neurotype and schedule. Be ready to take criticism: too often, work is set up as dictated by convention, rather than by calling on relevant experiences and possibility.

Make your teaching and scholarly materials multi-modal: produce formats that work for people with different physical conditions and ways of reading and communicating, sharing and contributing. Think about multiple ways to allow participation in your funding, reviewing, research and engagement. Let's see an end to patronizing objectification and assumptions about what we want and need: include disabled people in boards, teams and studies, and learn from us. We have had to become masters of invention. The pandemic has made the value of that clearer than ever.



# News in brief

## SCIENTISTS WARN: DON'T DRINK THE STAGNANT WATER

As some regions prepare to lift the lockdowns imposed in response to COVID-19, scientists who study water are worried about the potential for a secondary health crisis waiting in the dormant plumbing of offices, gyms, restaurants and schools, which could house increased levels of pathogens and heavy metals.

"There really isn't a lot of scientific awareness of these larger buildings," says Andrew Whelton, an environmental engineer at Purdue University in West Lafayette, Indiana. "Because of that, there's no guidance."

When Purdue's campus shut down in March, Whelton's laboratory began sampling water in several large buildings across the university.

Water quality in a complex building can vary by season, by time of day or even from room to room. So one-size-fits-all guidelines for flushing systems are unlikely to help, says Purdue environmental engineer Caitlin Proctor. The most common advice is to run all taps at full flow for a set amount of time – usually 5 or 10 minutes – to flush the stagnant water from the pipes.

But one building that the Purdue team studied required more than a full day to flush.



## DRUG REMDESIVIR AUTHORIZED TO TREAT CORONAVIRUS

One of the world's best hopes for treating COVID-19 – a compound called remdesivir – has been authorized as a therapy against the disease. On 1 May, the US Food and Drug Administration (FDA) granted an 'emergency-use authorization' for clinicians to use the drug, which is administered intravenously, in hospitals for people with severe COVID-19.

The authorization came after results from the largest and most rigorous clinical trial yet of the compound showed that the antiviral could shorten the time to recovery from coronavirus infection. Several smaller trials had reported mixed results in the past few weeks, but some had no placebo-controlled arm, making the results hard to interpret.

On 29 April, Anthony Fauci, director of the US National Institute of Allergy and Infectious Diseases (NIAID), said that the latest clinical trial, involving more than 1,000 people, showed that those taking remdesivir recovered in 11 days on average, compared with 15 days for those on a placebo. "Although a 31% improvement doesn't seem like a knockout 100%, it is a very important proof of concept," Fauci said. "What it has proven is that a drug can block this virus." He added that remdesivir would

become a standard treatment for COVID-19 in the United States.

Remdesivir, which was previously tested against Ebola, interferes with an enzyme that some viruses use to replicate; these include SARS-CoV-2, which is responsible for the current pandemic.

The latest trial results, which have not yet been published in a peer-reviewed journal, also showed that there were fewer deaths among participants who received the drug, Fauci said, but that trend was not statistically significant. However, the shortened recovery time was significant and beneficial enough for investigators to stop the trial early to ensure that participants receiving a placebo could access the drug.

The FDA's authorization is not a final drug approval, and can be revoked when the conditions required for emergency use are no longer in effect. US distribution of the drug will be government-controlled.

"It may not be the wonder drug that everyone's looking for, but if you can stop some patients from becoming critically ill, that's good enough," says virologist Stephen Griffin at the University of Leeds, UK. The drug's maker is Gilead Sciences in Foster City, California.

## HISTORICAL SPEND ON CORONAVIRUS RESEARCH WAS TINY

Before the COVID-19 pandemic, funding for coronavirus research made up just 0.5% of global infectious-diseases spending by public and philanthropic research organizations.

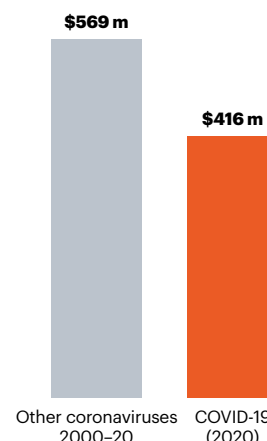
From 2000 to the start of this year, about US\$550 million was spent on coronavirus-related research, reveals an analysis by researchers at the University of Southampton, UK. By comparison, Ebola-related research got \$1.2 billion (1.1%).

Spending has risen to \$985 million since the current outbreak began. Some \$275 million of COVID-19 research funding is focused on vaccine development; \$40 million is focused on therapeutics and \$18 million on diagnostic tests. The figures cover spending by more than 1,000 funders worldwide.

The researchers note that the spending has generally been reactive – explaining spikes in 2004 and 2015, after outbreaks of the coronaviruses that cause the diseases SARS and MERS, respectively.

## CORONAVIRUS CASH

Public and philanthropic bodies have spent US\$985 million on coronavirus research since 2000. But before 2020, this constituted only 0.5% of total spending on infectious-disease research.





# News in focus



IVAN ROMANO/GETTY

Lockdown in Italy: scientists are working out what effects specific measures, such as social distancing, have in slowing the spread of COVID-19.

## WHOSE CORONAVIRUS STRATEGY WORKED BEST? SCIENTISTS HUNT MOST EFFECTIVE POLICIES

Researchers sift through data to compare nations' vastly different containment measures.

By Elizabeth Gibney

**H**ong Kong seems to have given the world a lesson in how to curb COVID-19 effectively. With a population of 7.5 million, it has reported just 4 deaths. Researchers studying its approach have found that swift surveillance, quarantine and social-distancing measures, such as school closures, helped to cut coronavirus transmission – measured by the average number of people each infected person infects, or  $R$  – to close to the critical level of 1 by early February. But the paper, published

in April, couldn't tease apart the effects of the various measures and behavioural changes happening at the same time (B. Cowling *et al. Lancet Public Health* 5, e279–e288; 2020).

Working out the effectiveness of the measures implemented worldwide to limit the coronavirus's spread is now one of scientists' most pressing questions. Researchers hope that, ultimately, they will be able to accurately predict how adding and removing control measures affects transmission rates and infection numbers. This information will be vital to governments as they design strategies to return life to normal, while keeping transmission low

to prevent second waves of infection. "This is not about the next epidemic. It's about 'what do we do now?'" says Rosalind Eggo, a mathematical modeller at the London School of Hygiene and Tropical Medicine (LSHTM).

Researchers are already working on models that use data from individual countries to understand the effect of various control measures. Models based on real data should be more nuanced than those that, at the start of the outbreak, predicted the effects of interventions mainly using assumptions. Combining data from around the world will allow scientists to compare countries' responses. And this should

## News in focus

allow them to design models that can make more accurate predictions about new phases of the pandemic and across nations.

But untangling cause and effect is extremely challenging, in part because circumstances differ in each country and because there is uncertainty over how well people adhere to control measures, cautions Eggo. “It’s really hard but it doesn’t mean we shouldn’t try.”

### Pulling together

Efforts to tackle these questions will soon get a boost from a database that brings together information on the hundreds of different interventions introduced worldwide. The platform, being prepared for the World Health Organization (WHO) by a team at the LSHTM, gathers data collected by ten groups already tracking interventions. These include researchers at the University of Oxford, UK, and at the Complexity Science Hub Vienna (CSH Vienna).

The database will standardize the information collected by the different teams and should be more comprehensive than anything an individual group could generate, says Chris Grundy, a data scientist behind the LSHTM project. Agencies such as the WHO routinely track control measures used in an outbreak, but for COVID-19 the picture is complicated by the pandemic’s speed and scale, says Grundy. The data set will be open, and improved in future releases, he says. Speed is of the essence: “Days make a difference right now.”

The trackers lay bare the vast range of policies deployed in different nations. The Vienna

team has captured details of around 170 interventions in 52 countries, ranging from small measures such as floor stickers that mark a 2-metre separation to major policies such as school closures. The team is also following some countries’ recent efforts to restart normal daily life. Oxford’s project, the COVID-19 Government Response Tracker, is monitoring 13 interventions in more than 100 countries. It compiles 7 of the 13 into a single ‘stringency’ index that captures the overall severity of each country’s response and allows for comparison between countries’ approaches (see ‘Pandemic protections’).

Already, scientists are analysing their data to explore differences in responses. The Vienna

**“If we don’t know what works, it’s going to be really difficult to decide what to do next.”**

team is looking for patterns, using methods that include clustering countries by how early in their epidemics they began introducing interventions and by the total number of restrictions introduced. In Europe, for example, algorithms group Sweden, the United Kingdom and the Netherlands together as countries that acted relatively slowly.

Meanwhile, Germany and Austria stand out as nations that adopted aggressive and early control strategies, compared with Italy, France

and Spain, which implemented similar measures, including lockdowns, but later in their epidemics. So far, Germany and Austria have, per capita, seen a fraction of the deaths from COVID-19 reported in these other countries.

Early findings from the Oxford team also suggest that poorer nations tended to bring in stricter measures than did richer countries, relative to the severity of their outbreaks. For example, the Caribbean nation of Haiti enforced lockdown on confirming its first case, whereas the United States waited until more than two weeks after its first death to issue stay-at-home orders. That might be because lower-income countries with less-developed health-care systems act more cautiously, says Anna Petherick, a public-policy researcher at Oxford. It could also reflect the fact that the outbreak reached these nations later, giving them longer to learn from others, she says.

### Patterns and predictions

Ultimately, researchers hope to use data from the LSHTM database to understand how effective these strategies were in limiting the outbreak. “We really need to evaluate those interventions in real time, so everybody can make real policies,” says Eggo, who was not involved in creating the database but plans to use it. “If we don’t know what works and we don’t know how much, it’s going to be really difficult to decide what to do next.” Eggo and her colleagues will use the data to test the accuracy of mathematical models – which use equations to describe the rate of transmission and mechanisms behind it – under varying intervention types and timing.

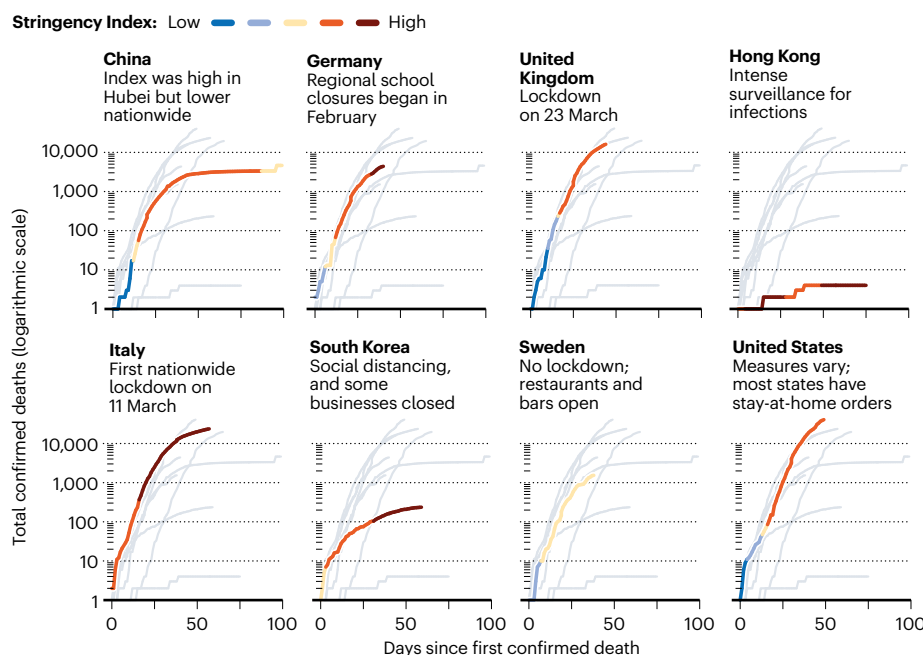
Ideally, researchers will be able to forecast how adding and removing interventions would change the number of infections over time. Policymakers could use such predictions, together with data on intensive-care capacity, to make decisions – on whether to reopen schools, for example – says Nils Haug, a mathematical physicist at CSH Vienna and the Medical University of Vienna.

Haug and his team of modellers are exploring which statistical approaches to use. Rather than directly determining the precise effect of each intervention, these methods can find ways to identify the measures that best predict infection rates. One approach involves using a machine-learning technique called a recurrent neural network to learn from patterns in the data and make predictions. Scientists can learn how important a given intervention is by looking at how predictions shift when they remove information about it from the network.

Without a vaccine or effective treatment, stopping transmission remains the only defence against COVID-19. Knowing the effects of each control measure is crucial to figuring out which ones can be safely altered or removed, says Petherick. “I think that would be a huge contribution.”

## PANDEMIC PROTECTIONS

Researchers have created a ‘stringency index’ that describes the overall severity of a country’s response to the coronavirus outbreak and allows responses to be compared. The index takes into account seven control measures, such as school closures and restrictions on people’s movements.



Confirmed deaths undercount true COVID-19 mortality. Stringency index developed by the Oxford COVID-19 Government Response Tracker. Data downloaded on 21 April; countries vary in day of most recent data update.

SOURCE: OXFORD CORONAVIRUS GOVERNMENT RESPONSE TRACKER (DATA); NATURE (CHARTS).





A coal-fired power plant in Colorado will not face stricter standards.

# FIVE WAYS TRUMP IS UNDERMINING ENVIRONMENTAL HEALTH

While the world focuses on a pandemic, the United States is adopting controversial environment policies.

By Jeff Tollefson

**T**he US Environmental Protection Agency (EPA) turns 50 this year, but scientists and environmentalists see little reason to celebrate. In the middle of a global pandemic, the agency's leadership is pressing forward with efforts to roll back environmental regulations and alter the way in which science is used to craft policy.

In the past month alone, the agency has dialled down regulations on automobile emissions and fuel efficiency; it has weakened rules on mercury and other pollutants emitted by power plants; and it has shied away from strengthening standards to reduce fine-particle air pollution.

"This is an extremely aggressive agenda," says Betsy Southerland, who spent more than three decades as an EPA official before retiring in protest against the current administration's policies in 2017.

## Two rollbacks on emissions

At the end of March, President Donald Trump's administration finalized a plan to scale back targets for automobile-emissions reductions from 5% per year to 1.5%, a change that the EPA acknowledges could result in an extra

867 million tonnes of carbon dioxide being released into the atmosphere by vehicles sold over the next decade.

In mid-April, the EPA issued another rule, targeting mercury-emissions standards for power plants first put in place under the previous president, Barack Obama. Although the agency left the existing limits in place, it adjusted how the rules' costs and benefits are calculated, weakening their economic justification. The original price tag reported for the 2011 regulation took into account health benefits from a reduction in particulate matter that would accompany cuts to mercury emissions.

Taking these out of the equation makes the rule seem more expensive, says David Spence, a political scientist and law scholar at the University of Texas at Austin. It also sets a precedent that could undermine the mercury rule and others.

## Inaction on particulates

Even more alarming, public-health experts say, was a decision on fine-particle pollution that EPA administrator Andrew Wheeler announced in mid-April. In that case, the EPA went against the advice of its own staff and many academic scientists by leaving the current standards in place – in spite of

evidence that reducing such pollution could save thousands of lives each year (Q. Di *et al.* *N. Engl. J. Med.* 376, 2513–2522; 2017).

In a report issued last September, EPA staff cited epidemiological and other evidence that would support cutting the maximum allowed average level of fine particulate matter from 12 micrograms per cubic metre of air to between 8 and 10.

The regulatory process that prevented that change was tipped towards the interests of polluters from the outset, with little to no independent scientific oversight, says Christopher Frey, an environmental engineer at North Carolina State University in Raleigh. Frey formerly chaired the EPA's scientific advisory committee on clean air, and was a member of a review panel for the issue that was disbanded in October 2018.

"Rather than focusing on protecting public health, EPA is on a misguided mission to protect the profits of regulated industries," Frey says. "But it's all based on a lot of misconceptions and assumptions, rather than facts or evidence."

## Two proposed rules

Looking ahead, it could become even more difficult to bring health data and other evidence to the policymaking table if the EPA moves forward with a pair of proposals that would alter how science is used and evaluated at the agency. First is a "transparency" rule that could restrict the use of public-health studies – including much of the epidemiological research that the agency has used to set particulate-pollution standards in the past.

A draft proposal states that if underlying data and models are not publicly available – often the case for private health-care data – the EPA could give them less weight or exclude them from consideration when setting standards and conducting scientific assessments.

"It would apply to pretty much all of EPA's major work," says Michael Halpern, deputy director of the Center for Science and Democracy at the Union of Concerned Scientists, an advocacy group in Cambridge, Massachusetts. Amendments to the proposal are open to public comment until 18 May.

A second proposal, currently pending review at the White House, would change how the agency evaluates the costs and benefits of environmental and public-health regulations, much as it did in its re-evaluation of the mercury-emissions standards.

Taken together, the cost-benefit guidance and the transparency rules could help the Trump administration to justify removing regulations, and could hamper regulatory efforts by future administrations. These and other EPA decisions will be challenged in court, but scientists and environmentalists say that provides little solace. "They are rolling back progress, and we are losing time," Halpern says.

# REFUGEE CAMPS RACE TO AVERT CORONAVIRUS CATASTROPHE

From Bangladesh to Greece, researchers and aid workers are taking steps to protect people.

By Nidhi Subbaraman

**T**here are 70 million refugees, displaced people and asylum seekers around the world. Close living quarters, significant underlying health problems and limited access to sanitation and medical care mean that COVID-19 – which has brought some of the best health systems in the world to their knees – poses an outsize threat to these communities.

Although there are some reports of refugees testing positive for the virus, as of mid-April there had been no known COVID-19 outbreaks in major refugee camps, according to advocates and responders contacted by *Nature*. But many aid groups fear that it is only a matter of time before the disease strikes. According to advocacy groups, host nations have been slow to enforce preventive measures. And experts fear that aid organizations will struggle to rally and respond.

## Models in Cox's Bazar

Nearly 600,000 Rohingya people now live in the Kutupalong-Balukhali Expansion Site in Cox's Bazar, Bangladesh, having fled persecution in Myanmar. The refugee camp is one of the world's biggest and most densely populated. And it is the first to be used in a model of COVID-19, says Paul Spiegel, director of the Center for Humanitarian Health at Johns

Hopkins University in Baltimore, Maryland.

Spiegel and his team projected outcomes at the camp using data, mainly from China, on age, case severity and case fatality rates for the outbreak (S. Truelove *et al.* Preprint at medRxiv <http://doi.org/ggqtq4>; 2020). The analysis has not been peer reviewed.

The group modelled outcomes in low-, moderate- and high-transmission scenarios. The worst-case scenario exhausts the encampment's medical capacity in just 58 days. When on-site hospitals are overwhelmed with COVID-19, there could be spikes in deaths from other diseases, such as malaria.

A spokesperson for the United Nations High Commissioner for Refugees says that the organization has used this model to guide its response, alongside partner organizations.

## A plan to move in Greece

Five Aegean islands off the coast of Greece host refugees and asylum seekers from the Middle East and south and central Asia before they can enter the country. Greece has facilities for about 6,000 people awaiting asylum decisions at these island entry points, but the camps have swelled to nearly 40,000 people. Many are living in rudimentary encampments among the olive groves, with limited access to running water and toilets.

"They are living in tents, makeshift camps and carton boxes. It's a highly overcrowded situation," says Apostolos Veizis, director of the medical operational support unit in Athens for aid organization Médecins Sans Frontières (MSF, also known as Doctors without Borders). Although other parts of the world are prioritizing social distancing and lockdowns to slow the spread of the coronavirus, the options are limited in a refugee camp. "There's not really an ability to self-isolate or quarantine within these communities," says Devon Cone, a senior advocate with Refugees International, a humanitarian organization based in Washington DC.

In response to the pandemic, MSF is prioritizing provision of water and sanitation facilities to the camps. Veizis says there is a plan to protect about 2,400 high-risk residents – those over 60 and those with chronic conditions – by moving them to mainland hotels. And in April, the European Union pledged €350 million (US\$377 million) to support refugees and asylum seekers, including those in Greece.

## Q&A

### Mission to trace



**Former World Bank president Jim Yong Kim is spearheading an ambitious US\$44-million project to control COVID-19 in the state of Massachusetts. The effort will involve a legion of contact tracers, who will phone people diagnosed with COVID-19 and their contacts, in the hope of isolating those infected and stopping the spread of coronavirus.**

#### What surprised you about the response to COVID-19 in the United States?

After the first cases of community transmission, in Washington state, folks were saying it's too late for containment. What surprised me is just how quickly we gave up on the standard shoe-leather-epidemiology approach to fighting epidemics that has been in place for hundreds of years.

#### What did you do?

I reached Charlie Baker, the governor of Massachusetts. And he made the decision to take a chance on a contact-tracing programme for the entire state. We're going to hire 1,000 people right away, and we'll have to spend more on isolation and quarantine. He had to make the decision to dip into the state's funding to do this.

#### Do you have any tips for how scientists can have their voices heard by policymakers?

I think it's very hard to convince others to do things that are complicated and labour-intensive, unless you've done it yourself, or at least started doing it.

#### Do you make a financial argument in your push to contain COVID-19?

I spent seven years at the World Bank. I can tell you it makes no sense to keep putting trillions of dollars into a stimulus package and not put hundreds of billions into the one thing that will stop the financial crisis: getting the virus under control.

We will probably get a vaccine, but we cannot stay in lockdown until that happens. The financial impact is just going to be so devastating that we have to figure out a different way.

#### Interview by Amy Maxmen

This interview has been edited for length and clarity.



An aid worker in Greece distributes masks.



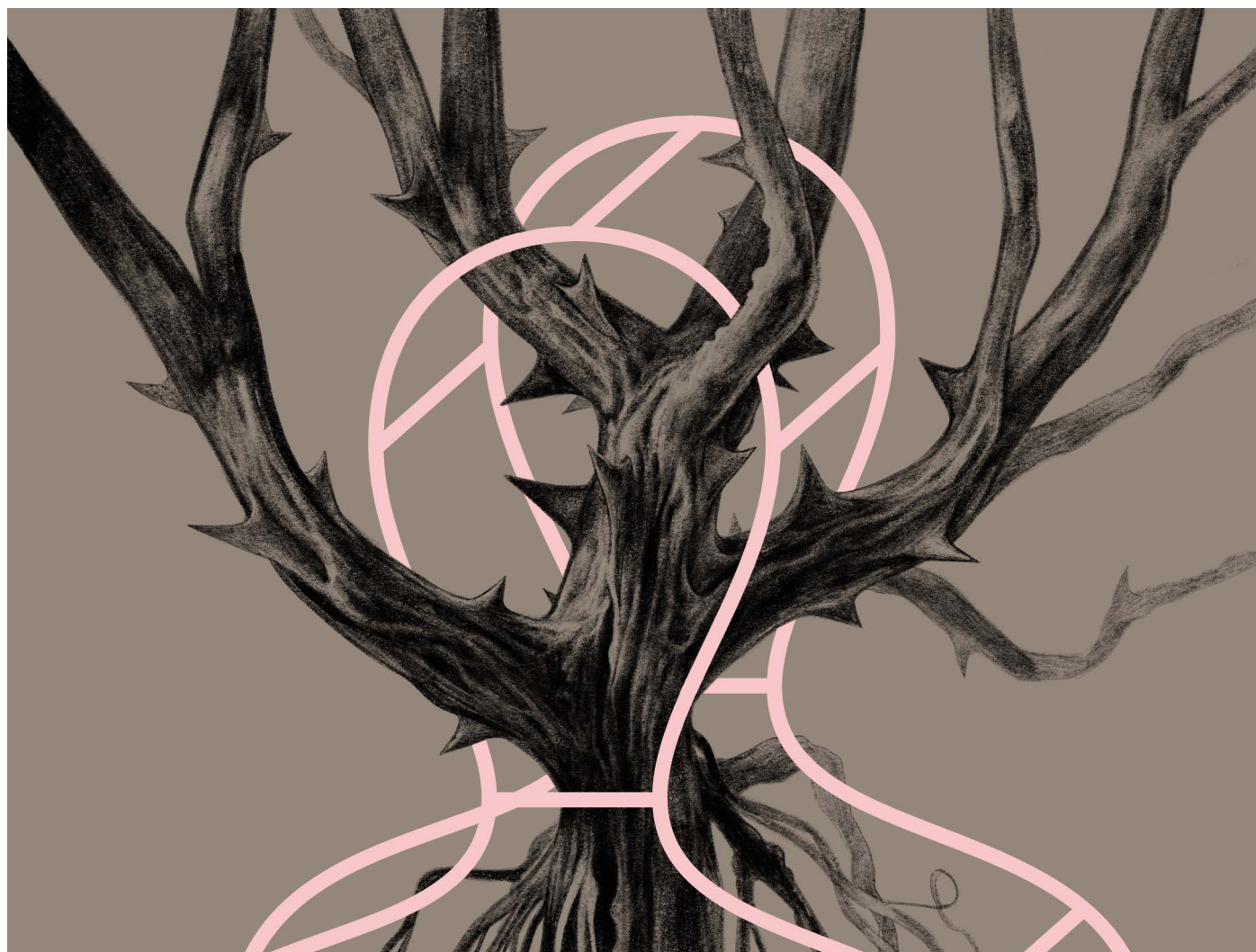


ILLUSTRATION BY PAWEŁ JONCA

# ROOTS OF MENTAL ILLNESS

Researchers are beginning to untangle the common biology that links supposedly distinct psychiatric conditions. **By Michael Marshall**

In 2018, psychiatrist Oleguer Plana-Ripoll was wrestling with a puzzling fact about mental disorders. He knew that many individuals have multiple conditions — anxiety and depression, say, or schizophrenia and bipolar disorder. He wanted to know how common it was to have more than one diagnosis, so he got his hands on a database containing the medical details of around 5.9 million Danish citizens.

He was taken aback by what he found. Every single mental disorder predisposed the patient to every other mental disorder — no matter how distinct the symptoms<sup>1</sup>. “We knew

that comorbidity was important, but we didn’t expect to find associations for all pairs,” says Plana-Ripoll, who is based at Aarhus University in Denmark.

The study tackles a fundamental question that has bothered researchers for more than a century. What are the roots of mental illness?

In the hope of finding an answer, scientists have piled up an enormous amount of data over the past decade, through studies of genes, brain activity and neuroanatomy. They have found evidence that many of the same genes underlie seemingly distinct disorders, such as schizophrenia and autism, and that changes in

the brain’s decision-making systems could be involved in many conditions.

Researchers are also drastically rethinking theories of how our brains go wrong. The idea that mental illness can be classified into distinct, discrete categories such as ‘anxiety’ or ‘psychosis’ has been disproved to a large extent. Instead, disorders shade into each other, and there are no hard dividing lines — as Plana-Ripoll’s study so clearly demonstrated.

Now, researchers are trying to understand the biology that underlies this spectrum of psychopathology.

They have a few theories. Perhaps there

## Feature

are several dimensions of mental illness – so, depending on how a person scores on each dimension, they might be more prone to some disorders than to others. An alternative, more radical idea is that there is a single factor that makes people prone to mental illness in general: which disorder they develop is then determined by other factors. Both ideas are being taken seriously, although the concept of multiple dimensions is more widely accepted by researchers.

The details are still fuzzy, but most psychiatrists agree that one thing is clear: the old system of categorizing mental disorders into neat boxes does not work. They are also hopeful that, in the long run, replacing this framework with one that is grounded in biology will lead to new drugs and treatments. Researchers aim to reveal, for instance, the key genes, brain regions and neurological processes involved in psychopathology, and target them with therapies. Although it might take a while to get there, says Steven Hyman of the Broad Institute of MIT and Harvard in Cambridge, Massachusetts, “I am long-term optimistic if the field really does its work.”

### A smorgasbord of disorders

The most immediate challenge is working out how to diagnose people. Since the 1950s, psychiatrists have used an exhaustive volume called the *Diagnostic and Statistical Manual of Mental Disorders*, currently in its fifth edition. It lists all the recognized disorders, from autism and obsessive-compulsive disorder to depression, anxiety and schizophrenia. Each is defined by symptoms. The inherent assumption is that each disorder is distinct, and arises for different reasons.

However, even before the *DSM-5* was published in 2013, many researchers argued that this approach was flawed. “Any clinician could have told you that patients had not read the *DSM* and didn’t conform to the *DSM*,” says Hyman, who helped to draft the manual’s fifth edition.

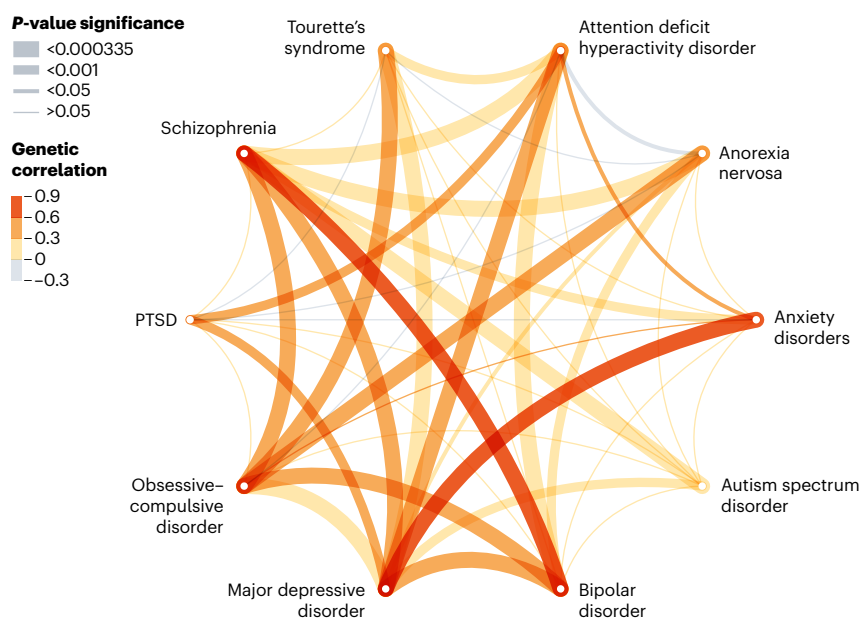
Few patients fit into each neat set of criteria. Instead, people often have a mix of symptoms from different disorders. Even if someone has a fairly clear diagnosis of depression, they often have symptoms of another disorder such as anxiety. “If you have one disorder, you’re much more likely to have another,” says Ted Satterthwaite, a neuropsychiatrist at the University of Pennsylvania in Philadelphia.

This implies that the way clinicians have partitioned mental disorders is wrong. Psychiatrists have tried to solve this by splitting disorders into ever-finer subtypes. “If you look at the way the *DSM* has evolved over time, the book gets thicker and thicker,” says Satterthwaite. But the problem persists – the subtypes are still a poor reflection of the clusters of symptoms that many patients have.

As a result, the world’s largest funder of mental-health science, the US National Institute of Mental Health, changed the way it funded research. Beginning in 2011, it began

### MENTAL MAP

Similar genetic variants seem to underlie a number of psychiatric disorders. In one study of 200,000 people<sup>8</sup>, schizophrenia was significantly correlated with most other disorders. By contrast, some disorders such as post-traumatic stress disorder (PTSD) showed only weak correlations to other conditions.



demanding more studies of the biological basis of disorders, instead of their symptoms, under a programme called the Research Domain Criteria. There has since been an explosion of research into the biological basis of psychopathology, with studies focusing on genetics and neuroanatomy, among other fields. But if researchers hoped to demystify psychopathology, they still have a long way to go: the key finding has been just how complex psychopathology really is.

### Controversial clusters

Clinically, the evidence that symptoms cut across disorders – or that people frequently have more than one disorder – has only grown stronger. For this reason, although individual symptoms such as mood alterations or impairments in reasoning can be diagnosed reliably, assigning patients to an overall diagnosis such as ‘bipolar disorder’ is difficult.

Even seemingly separate disorders are linked. In 2008, geneticist Angelica Ronald, then at King’s College London Institute of Psychiatry, and her colleagues found that autism and attention deficit hyperactivity disorder (ADHD) overlapped. “At the time, you weren’t allowed to be diagnosed with both conditions,” says Ronald; this was because of a rule in an earlier version of the *DSM*. But she and her team found that traits for autism and ADHD were strongly correlated, and partially under genetic control<sup>2</sup>.

Furthermore, there seem to be clusters of symptoms that cross the boundaries of disorders. A 2018 study<sup>3</sup> examined people who had been diagnosed with either major depression, panic disorder or post-traumatic stress disorder (PTSD). The volunteers were assessed on the basis of their symptoms, cognitive performance

and brain activity. The researchers found that the participants fell into six groups, characterized by distinct moods such as ‘tension’ and ‘melancholia’. The groups cut across the three diagnostic categories as if they were not there.

Many now agree that the diagnostic categories are wrong. The question is, with biology as their guide, what should psychiatric diagnosis and treatment look like instead?

### Multiple dimensions

One prominent model is that there are a number of neuropsychological traits or ‘dimensions’ that vary in every person. Each trait determines our susceptibility to certain kinds of disorder. For example, someone might be prone to mood disorders such as anxiety, but not to thought disorders such as schizophrenia.

This is similar to the way psychologists think about personality. In one model, five personality traits, such as conscientiousness and neuroticism, describe most of the variation in human personalities.

Some psychiatrists are already trying to reimagine their discipline with dimensions in mind. In the early 2010s, there was a push to eliminate disorder categories from the *DSM-5* in favour of a ‘dimensional’ approach based on individual symptoms. However, this attempt failed – partly because health-care funding and patient care has been built up around the *DSM*’s categories. However, other catalogues of disorders have shifted towards dimensionality. In 2019, the World Health Assembly endorsed the latest International Classification of Diseases (called ICD-11), in which some psychopathologies were newly broken down using dimensional symptoms rather than categories.

SOURCE: REF. 8



The challenge for the dimensionality hypothesis is obvious: how many dimensions are there, and what are they? Satterthwaite calls this “a very large problem”.

One popular theory<sup>4</sup>, supported by many studies over the past decade, argues for just two dimensions. The first includes all ‘internalizing’ disorders, such as depression, in which the primary symptoms affect a person’s internal state. This is contrasted with ‘externalizing’ disorders, such as hyperactivity and antisocial behaviour, in which a person’s response to the world is affected. If someone has been diagnosed with two or more disorders, studies suggest these are likely to be from the same category.

But studies combining large amounts of brain-imaging data with machine learning have turned up different numbers – even in studies done by the same lab. Last year, Satterthwaite and his group published a study<sup>5</sup> of 1,141 young people who had internalizing symptoms, and found they could be split into two groups on the basis of their brain structure and function. In 2018, Satterthwaite led a similar study<sup>6</sup> and identified four dimensions, each associated with a distinct pattern of brain connectivity.

Ultimately, a future version of the *DSM* could have chapters devoted to each dimension, says Hyman. These could list the disorders that cluster within each, as well as their symptoms and any biomarkers derived from the underlying physiology and genetics. Two people who had similar symptoms but different sets of mutations or neuroanatomical alterations could then be diagnosed and treated differently.

## In the genes

One pillar of this future approach is a better understanding of the genetics of mental illness. In the past decade, studies of psychopathological genetics have become large enough to draw robust conclusions.

The studies reveal that no individual gene contributes much to the risk of a psychopathology; instead, hundreds of genes each have a small effect. A 2009 study<sup>7</sup> found that thousands of gene variants were risk factors for schizophrenia. Many were also associated with bipolar disorder, suggesting that some genes contribute to both disorders.

This is not to say that the same genes are involved in all brain disorders: far from it. A team led by geneticist Benjamin Neale at Massachusetts General Hospital in Boston and psychiatrist Aiden Corvin at Trinity College Dublin found in 2018 that neurological disorders such as epilepsy and multiple sclerosis are genetically distinct from psychiatric disorders such as schizophrenia and depression<sup>8</sup> (see ‘Mental map’).

These studies all looked at common variants, which are the easiest to detect. Some recent studies focused instead on extremely rare variants, which do suggest genetic differences between disorders. A study of more

than 12,000 people<sup>9</sup> found that individuals with schizophrenia had an unusually high rate of ultra-rare mutations – and that these were often unique to one individual.

The result is a mess. It is difficult to predict which risk factors cut across conditions. “Some of them are quite broadly shared across psychopathology,” says Neale, “whereas some are a bit more specific to one or a handful of forms of psychopathology.”

## The *p* factor

Some psychiatrists have put forward a radical hypothesis that they hope will allow them to make sense of the chaos. If disorders share symptoms, or co-occur, and if many genes are implicated in multiple disorders, then maybe there is a single factor that predisposes people to psychopathology.

**“If you have one disorder, you’re much more likely to have another.”**

The idea was first proposed in 2012 by public-health specialist Benjamin Lahey at the University of Chicago in Illinois<sup>10</sup>. Lahey and his colleagues studied symptoms in 11 disorders. They used statistics to examine whether the pattern could best be explained by three distinct dimensions, or by those three together with a ‘general’ predisposition. The model worked better if the general factor was included.

The following year, the hypothesis received more support – and a catchy name – from husband-and-wife psychologists Avshalom Caspi and Terrie Moffitt at Duke University in Durham, North Carolina. They used data from a long-term study of 1,037 people and found that most of the variation in symptoms could be explained by a single factor<sup>11</sup>. Caspi and Moffitt called this the ‘*p* factor’. Since 2013, multiple studies have replicated their core finding.

Caspi and Moffitt were clear that the *p* factor could not explain everything, and made no guesses about its underlying biology, speculating only that a set of genes might mediate it. Others have proposed that the *p* factor is a general predisposition to psychopathology, but that other factors – stressful experiences, or other gene alterations – nudge a person towards different symptoms<sup>12</sup>. But if it is real, it has a startling implication: there could be a single therapeutic target for psychiatric disorders.

There are already hints that generalized treatments could work just as well as targeted therapies. A 2017 study<sup>13</sup> randomly assigned people with anxiety disorders, such as panic disorder or obsessive-compulsive disorder, to receive either a therapy for their specific disorder or a generalized approach. Both therapies worked equally well.

Finding a physiological basis for the *p* factor

would be the first step towards therapies based on it, but only in the past few years have researchers found hints of it in genetic and neuroanatomical data. One study<sup>14</sup> of the genetics of psychopathology in a UK population, for instance, identified a ‘genetic *p* factor’ – a set of genes in which there were variations that contributed to the risk of psychopathology.

Meanwhile, other groups have searched for a neuroanatomical change that occurs in multiple psychopathologies. The results are intriguing, but contradictory.

One study<sup>15</sup> of six psychopathologies found that the brain’s grey matter shrank in three regions involved in processing emotions: the dorsal anterior cingulate, right insula and left insula. But subsequent studies by Adrienne Romer, a clinical psychologist now at Harvard Medical School and McLean Hospital in Belmont, Massachusetts, identified a totally different trio of regions with roles that include managing basic bodily functions and movement<sup>16</sup> – the pons, cerebellum and part of the cortex. One key to making sense of this might be to focus on the brain’s executive function: the ability to regulate behaviour by planning, paying attention and resisting temptation, which relies on many brain regions. Romer and Satterthwaite have independently found disruptions in executive function in a range of psychopathologies<sup>17,18</sup> – the suspicion being that these disruptions could underlie the *p* factor.

Most scientists agree that what is needed is more data, and many remain unconvinced by such simple explanations. “I’m a little less certain that that’s how it’s going to play out,” says Neale. At the genetic level at least, he says, there are many disorders, such as PTSD and generalized anxiety disorder, that remain poorly understood.

All such sweeping hypotheses are premature, says Hyman. “I think it’s a time for much more empirical research rather than grand theorization.”

**Michael Marshall** is a freelance writer based in Devon, UK.

1. Plana-Ripoll, O. et al. *JAMA Psychiatry* **76**, 259–270 (2019).
2. Ronald, A., Simonoff, E., Kuntsi, J., Asherson, P. & Plomin, R. *J. Child Psychol. Psychiatry* **49**, 535–542 (2008).
3. Grisanzio, K. A. et al. *JAMA Psychiatry* **75**, 201–209 (2018).
4. Krueger, R. F. & Eaton, N. R. *World Psychiatry* **14**, 27–29 (2015).
5. Kaczkurkin, A. N. et al. *Biol. Psychiatry* **87**, 473–482 (2020).
6. Xia, C. H. et al. *Nature Commun.* **9**, 3003 (2018).
7. The International Schizophrenia Consortium. *Nature* **460**, 748–752 (2009).
8. The Brainstorm Consortium et al. *Science* **360**, eaap8757 (2018).
9. Genovese, G. et al. *Nature Neurosci.* **19**, 1433–1441 (2016).
10. Lahey, B. B. et al. *J. Abnorm. Psychol.* **121**, 971–977 (2012).
11. Caspi, A. et al. *Clin. Psychol. Sci.* **2**, 119–137 (2014).
12. Lahey, B. B., Krueger, R. F., Rathouz, P. J., Waldman, I. D. & Zald, D. H. *Psychol. Bull.* **143**, 142–186 (2017).
13. Barlow, D. H. et al. *JAMA Psychiatry* **74**, 875–884 (2017).
14. Selzam, S., Coleman, J. R. I., Caspi, A., Moffitt, T. E. & Plomin, R. *Transl. Psychiatry* **8**, 205 (2018).
15. Goodkind, M. et al. *JAMA Psychiatry* **72**, 305–315 (2015).
16. Romer, A. L. et al. *Mol. Psychiatry* **23**, 1084–1090 (2017).
17. Elliott, M. L. et al. *Biol. Psychiatry* **84**, 452–459 (2018).
18. Shanmugan, S. et al. *Am. J. Psychiatry* **173**, 517–526 (2016).

# PROFILE OF A

Scientists are quickly piecing together how the new coronavirus operates, where it came from and what it might do next – but pressing questions remain. **By David Cyranoski**

**I**n 1912, German veterinarians puzzled over the case of a feverish cat with an enormously swollen belly. That is now thought to be the first reported example of the debilitating power of a coronavirus. Veterinarians didn't know it at the time, but coronaviruses were also giving chickens bronchitis, and pigs an intestinal disease that killed almost every piglet under two weeks old.

The link between these pathogens remained hidden until the 1960s, when researchers in the United Kingdom and the United States isolated two viruses with crown-like structures causing common colds in humans. Scientists soon noticed that the viruses identified in sick animals had the same bristly structure, studded with spiky protein protrusions. Under electron microscopes, these viruses resembled the solar corona, which led researchers in 1968 to coin the term coronaviruses for the entire group.

It was a family of dynamic killers: dog coronaviruses could harm cats, the cat coronavirus could ravage pig intestines. Researchers thought that coronaviruses caused only mild symptoms in humans, until the outbreak of severe acute respiratory syndrome (SARS) in 2003 revealed how easily these versatile

viruses could kill people.

Now, as the death toll from the COVID-19 pandemic surges, researchers are scrambling to uncover as much as possible about the biology of the latest coronavirus, named SARS-CoV-2. A profile of the killer is already emerging. Scientists are learning that the virus has evolved an array of adaptations that make it much more lethal than the other coronaviruses humanity has met so far. Unlike close relatives, SARS-CoV-2 can readily attack human cells at multiple points, with the lungs and the throat being the main targets. Once inside the body, the virus makes use of a diverse arsenal of dangerous molecules. And genetic evidence suggests that it has been hiding out in nature possibly for decades.

## Bad family

But there are many crucial unknowns about this virus, including how exactly it kills, whether it will evolve into something more – or less – lethal and what it can reveal about the next outbreak from the coronavirus family.

“There will be more, either out there already or in the making,” says Andrew Rambaut, who studies viral evolution at the University of Edinburgh, UK.

Of the viruses that attack humans, coronaviruses are big. At 125 nanometres in diameter, they are also relatively large for the viruses that use RNA to replicate, the group that accounts for most newly emerging diseases. But coronaviruses really stand out for their genomes. With 30,000 genetic bases, coronaviruses have the largest genomes of all RNA viruses. Their genomes are more than three times as big as those of HIV and hepatitis C, and more than twice influenza's.

Coronaviruses are also one of the few RNA viruses with a genomic proofreading mechanism – which keeps the virus from accumulating mutations that could weaken it. That ability might be why common antivirals such as ribavirin, which can thwart viruses such as hepatitis C, have failed to subdue SARS-CoV-2. The drugs weaken viruses by inducing mutations. But in the coronaviruses, the proofreader can weed out those changes.

Mutations can have their advantages for viruses. Influenza mutates up to three times more often than coronaviruses do, a pace that enables it to evolve quickly and sidestep vaccines. But coronaviruses have a special trick that gives them a deadly dynamism: they frequently recombine, swapping chunks of their



# KILLER VIRUS



ILLUSTRATION BY FABIO BUONOCORE

## Feature

RNA with other coronaviruses. Typically, this is a meaningless trading of like parts between like viruses. But when two distant coronavirus relatives end up in the same cell, recombination can lead to formidable versions that infect new cell types and jump to other species, says Rambaut.

Recombination happens often in bats, which carry 61 viruses known to infect humans; some species harbour as many as 12 (ref. 1). In most cases, the viruses don't harm the bats, and there are several theories about why bats' immune systems can cope with these invaders. A paper published in February argues that bat cells infected by viruses rapidly release a signal that makes them able to host the virus without killing it<sup>2</sup>.

Estimates for the birth of the first coronavirus vary widely, from 10,000 years ago to 300 million years ago. Scientists are now aware of dozens of strains<sup>3</sup>, seven of which infect humans. Among the four that cause common colds, two (OC43 and HKU1) came from rodents, and the other two (229E and NL63) from bats. The three that cause severe disease – SARS-CoV (the cause of SARS), Middle East respiratory syndrome MERS-CoV and SARS-CoV-2 – all came from bats. But scientists think there is usually an intermediary – an animal infected by the bats that carries the virus into humans. With SARS, the intermediary is thought to be civet cats, which are sold in live-animal markets in China.

The origin of SARS-CoV-2 is still an open question. The virus shares 96% of its genetic material with a virus found in a bat in a cave in Yunnan, China<sup>4</sup> – a convincing argument that it came

from bats, say researchers. But there's a crucial difference. The spike proteins of coronaviruses have a unit called a receptor-binding domain, which is central to their success in entering human cells. The SARS-CoV-2 binding domain is particularly efficient, and it differs in important ways from that of the Yunnan bat virus, which seems not to infect people<sup>5</sup>.

Complicating matters, a scaly anteater called the pangolin showed up with a coronavirus that had a receptor-binding domain almost identical to the human version. But the rest of the coronavirus was only 90% genetically similar, so some researchers suspect the pangolin was not the intermediary<sup>5</sup>. The fact that both mutations and recombinations are at work complicates efforts to draw a family tree.

But studies released over the past few months, which have yet to be peer-reviewed, suggest that SARS-CoV-2 – or a very similar ancestor – has been hiding in some animal for decades. According to a paper posted online in March<sup>6</sup>, the coronavirus lineage leading to SARS-CoV-2 split more than 140 years ago from the closely related one seen today in pangolins. Then, sometime in the past 40–70 years, the ancestors of SARS-CoV-2 separated from the bat version, which subsequently lost the effective receptor binding domain that was present in its ancestors (and remains in SARS-CoV-2). A study published on 21 April came up with very similar findings using a different dating method<sup>7</sup>.

These results suggest a long family history, with many coronavirus branches in bats and possibly pangolins carrying the same deadly receptor binding domain as SARS-CoV-2, including some that might have similar

abilities to cause a pandemic, says Rasmus Nielsen, an evolutionary biologist at the University of California, Berkeley, and co-author of the second study. "There is a need for continued surveillance and increased vigilance towards the emergence of new viral strains by zoonotic transfer," he says.

### Two open doors

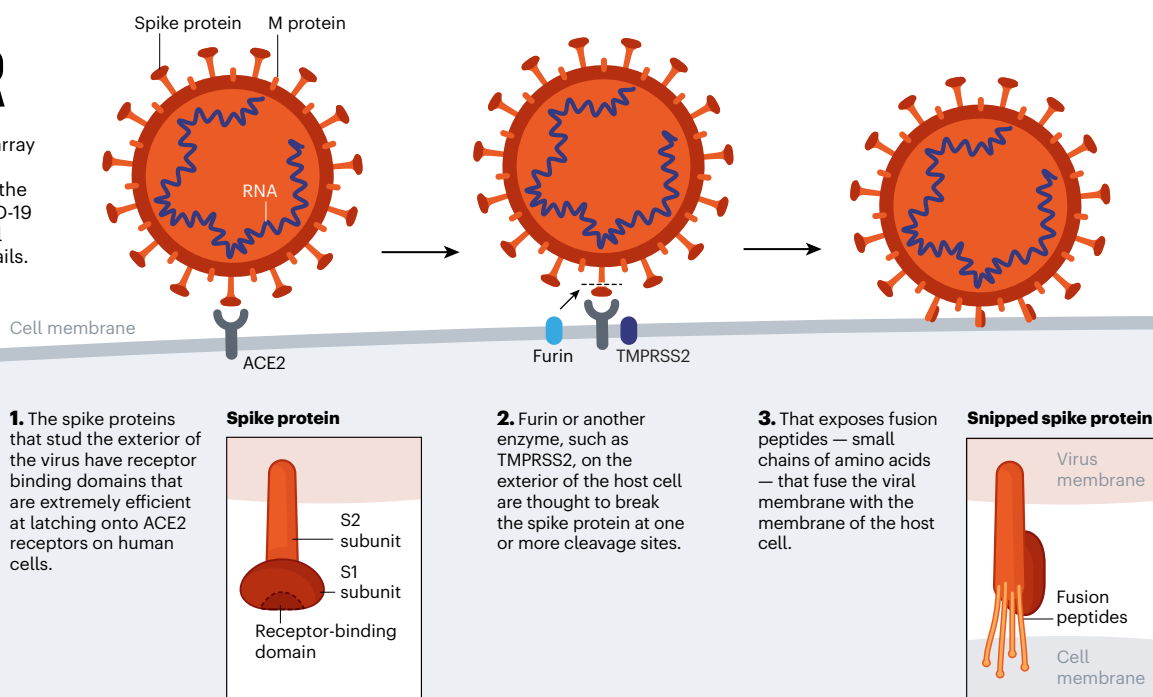
Although the known human coronaviruses can infect many cell types, they all mainly cause respiratory infections. The difference is that the four that cause common colds easily attack the upper respiratory tract, whereas MERS-CoV and SARS-CoV have more difficulty gaining a hold there, but are more successful at infecting cells in the lungs.

SARS-CoV-2, unfortunately, can do both very efficiently. That gives it two places to get a foothold, says Shu-Yuan Xiao, a pathologist at the University of Chicago, Illinois. A neighbour's cough that sends ten viral particles your way might be enough to start an infection in your throat, but the hair-like cilia found there are likely to do their job and clear the invaders. If the neighbour is closer and coughs 100 particles towards you, the virus might be able get all the way down to the lungs, says Xiao.

These varying capacities might explain why people with COVID-19 have such different experiences. The virus can start in the throat or nose, producing a cough and disrupting taste and smell, and then end there. Or it might work its way down to the lungs and debilitate that organ. How it gets down there, whether it moves cell by cell or somehow gets washed down, is not known, says Stanley Perlman, an

## DEADLY INVADER

Research suggests the SARS-CoV-2 virus has an array of adaptations that help it break into human cells – the first step in causing COVID-19 disease. Scientists are still debating many of the details.





immunologist at the University of Iowa in Iowa City who studies coronaviruses.

Clemens-Martin Wendtner, an infectious-disease physician at the Munich Clinic Schwabing in Germany, says it could be a problem with the immune system that lets the virus sneak down into the lungs. Most infected people create neutralizing antibodies that are tailored by the immune system to bind with the virus and block it from entering a cell. But some people seem unable to make them, says Wendtner. That might be why some recover after a week of mild symptoms, whereas others get hit with late-onset lung disease. But the virus can also bypass the throat cells and go straight down into the lungs. Then patients might get pneumonia without the usual mild symptoms such as a cough or low-grade fever that would otherwise come first, says Wendtner. Having these two infection points means that SARS-CoV-2 can mix the transmissibility of the common cold coronaviruses with the lethality of MERS-CoV and SARS-CoV. "It is an unfortunate and dangerous combination of this coronavirus strain," he says.

The virus's ability to infect and actively reproduce in the upper respiratory tract was something of a surprise, given that its close genetic relative, SARS-CoV, lacks that ability. Last month, Wendtner published results<sup>8</sup> of experiments in which his team was able to culture virus from the throats of nine people with COVID-19, showing that the virus is actively reproducing and infectious there. That explains a crucial difference between the close relatives. SARS-CoV-2 can shed viral particles from the throat into saliva even before



**WHEN I SAW SARS-COV-2 HAD THAT CLEAVAGE SITE, I DID NOT SLEEP VERY WELL THAT NIGHT."**

symptoms start, and these can then pass easily from person to person. SARS-CoV was much less effective at making that jump, passing only when symptoms were full-blown, making it easier to contain.

These differences have led to some confusion about the lethality of SARS-CoV-2. Some experts and media reports describe it as less deadly than SARS-CoV because it kills about 1% of the people it infects, whereas SARS-CoV killed at roughly ten times that rate. But Perlman says that's the wrong way to look at it. SARS-CoV-2 is much better at infecting people, but many of the infections don't progress to the lungs. "Once it gets down in the lungs, it's probably just as deadly," he says.

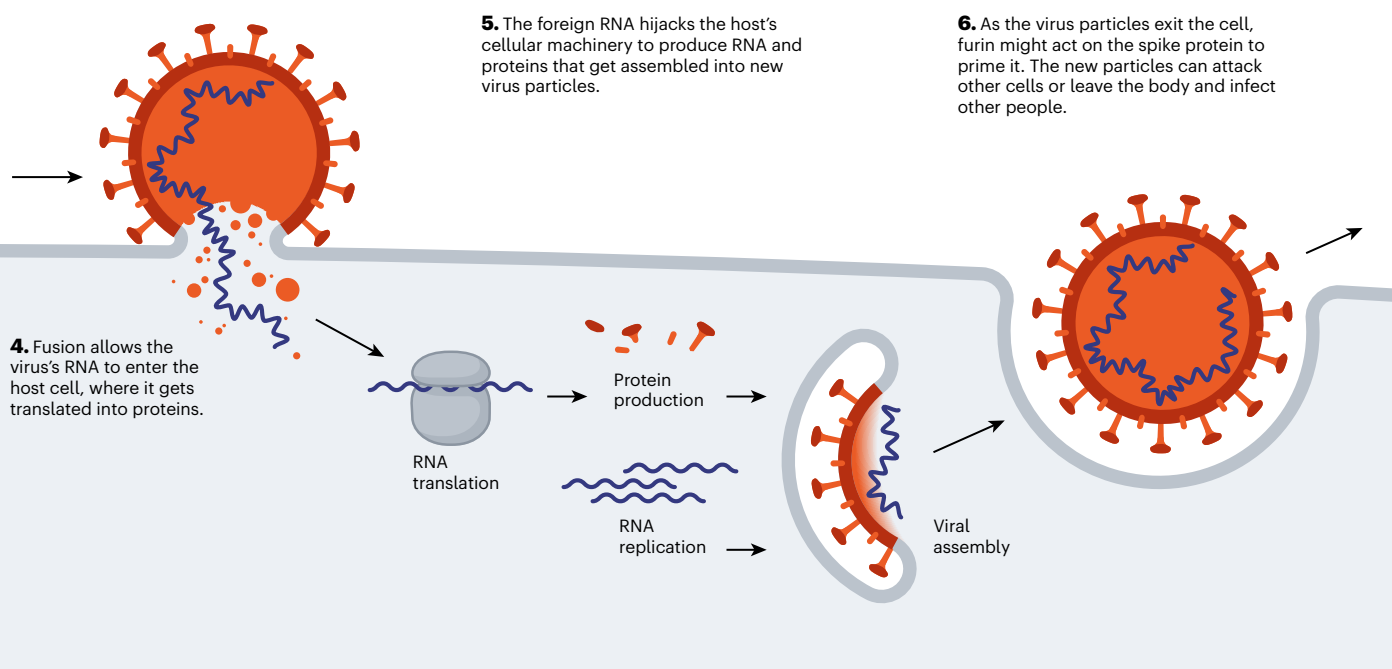
What it does when it gets down to the lungs is similar in some respects to what respiratory viruses do, although much remains unknown. Like SARS-CoV and influenza, it infects and destroys the alveoli, the tiny sacs in the lungs that shuttle oxygen into the bloodstream. As the cellular barrier dividing these sacs from blood vessels break down, liquid from

the vessels leaks in, blocking oxygen from getting to the blood. Other cells, including white blood cells, plug up the airway further. A robust immune response will clear all this out in some patients, but overreaction of the immune system can make the tissue damage worse. If the inflammation and tissue damage are too severe, the lungs never recover and the person dies or is left with scarred lungs, says Xiao. "From a pathological point of view, we don't see a lot of uniqueness here."

And as with SARS-CoV, MERS-CoV and animal coronaviruses, the damage doesn't stop with the lungs. A SARS-CoV-2 infection can trigger an excessive immune response known as a cytokine storm, which can lead to multiple organ failure and death. The virus can also infect the intestines, the heart, the blood, sperm (as can MERS-CoV), the eye and possibly the brain. Damage to the kidney, liver and spleen observed in people with COVID-19 suggests that the virus can be carried in the blood and infect various organs or tissues, says Guan Wei-jie, a pulmonologist at the Guangzhou Institute of Respiratory Health at Guangzhou Medical University, China, an institution lauded for its role in combating SARS and COVID-19. The virus might be able to infect various organs or tissues wherever the blood supply reaches, says Guan.

But although genetic material from the virus is showing up in these various tissues, it is not yet clear whether the damage there is being done by the virus or by a cytokine storm, says Wendtner. "Autopsies are under way in our centre. More data will come soon," he says.

Whether it infects the throat or the lungs,



SARS-CoV-2 breaches the protective membrane of host cells using its spike proteins (see ‘Deadly invader’). First, the protein’s receptor-binding domain latches on to a receptor called ACE2, which sits on the surface of the host cell. ACE2 is expressed throughout the body on the lining of the arteries and veins that course through all organs, but it is particularly dense on the cells lining the alveoli and small intestines.

Although the exact mechanisms remain unknown, evidence suggests that after the virus attaches itself, the host cell snips the spike protein at one of its dedicated ‘cleavage sites’, exposing fusion peptides – small chains of amino acids that help to pry open the host cell’s membrane so that the virus’s membrane can merge with it. Once the invader’s genetic material gets inside the cell, the virus commandeers the host’s molecular machinery to produce new viral particles. Then, those progeny exit the cell to go and infect others.

## Power spikes

SARS-CoV-2 is uniquely equipped for forcing entry into cells. Both SARS-CoV and SARS-CoV-2 bind with ACE2, but the receptor-binding domain of SARS-CoV-2 is a particularly snug fit. It is 10–20 times more likely to bind ACE2 than is SARS-CoV<sup>9</sup>. Wendtner says that SARS-CoV-2 is so good at infecting the upper respiratory tract that there might even be a second receptor that the virus could use to launch its attack.

Even more troubling is the fact that SARS-CoV-2 seems to make use of the enzyme furin from the host to cleave the viral spike protein. This is worrying, researchers say, because furin is abundant in the respiratory tract and found throughout the body. It is used by other formidable viruses, including HIV, influenza, dengue and Ebola to enter cells. By contrast, the cleavage molecules used by SARS-CoV are much less common and not as effective.

Scientists think that the involvement of furin could explain why SARS-CoV-2 is so good at jumping from cell to cell, person to person and possibly animal to human. Robert Garry, a virologist at Tulane University in New Orleans, Louisiana, estimates that it gives SARS-CoV-2 a 100–1,000 times greater chance than SARS-CoV of getting deep into the lungs. “When I saw SARS-CoV-2 had that cleavage site, I did not sleep very well that night,” he says.

The mystery is where the genetic instructions for this particular cleavage site came from. Although the virus probably gained them through recombination, this particular set-up has never been found in any other coronavirus in any species. Pinning down its origin might be the last piece in the puzzle that will determine which animal was the stepping stone that allowed the virus to reach humans.

Some researchers hope that the virus will weaken over time through a series of

mutations that adapt it to persist in humans. By this logic, it would become less deadly and have more chances to spread. But researchers have not yet found any sign of such weakening, probably because of the virus’s efficient genetic repair mechanism. “The genome of COVID-19 virus is very stable, and I don’t see any change of pathogenicity that is caused by virus mutation,” says Guo Deyin, who researches coronaviruses at Sun Yat-sen University in Guangzhou.

Rambaut, too, doubts that the virus will become milder over time and spare its host. “It doesn’t work that way,” he says. As long as it can successfully infect new cells, reproduce and transmit to new ones, it doesn’t matter whether it harms the host, he says.

**BY FAR THE MOST LIKELY SCENARIO IS THAT THE VIRUS WILL CONTINUE TO SPREAD.”**

But others think there is a chance for a better outcome. It might give people antibodies that will offer at least partial protection, says Klaus Stöhr, who headed the World Health Organization’s SARS research and epidemiology division. Stöhr says that immunity will not be perfect – people who are reinfected will still develop minor symptoms, the way they do now from the common cold, and there will be rare examples of severe disease. But the virus’s proofreading mechanism means it will not mutate quickly, and people who were infected will retain robust protection, he says.

“By far the most likely scenario is that the virus will continue to spread and infect most of the world population in a relatively short period of time,” says Stöhr, meaning one to two years. “Afterwards, the virus will continue to spread in the human population, likely forever.” Like the four generally mild human coronaviruses, SARS-CoV-2 would then circulate constantly and cause mainly mild upper respiratory tract infections, says Stöhr. For that reason, he adds, vaccines won’t be necessary.

Some previous studies support this argument. One<sup>10</sup> showed that when people were inoculated with the common-cold coronavirus 229E, their antibody levels peaked two weeks later and were only slightly raised after a year. That did not prevent infections a year later, but subsequent infections led to few, if any, symptoms and a shorter period of viral shedding.

The OC43 coronavirus offers a model for

where this pandemic might go. That virus also gives humans common colds, but genetic research from the University of Leuven in Belgium suggests that OC43 might have been a killer in the past<sup>11</sup>. That study indicates that OC43 spilled over to humans in around 1890 from cows, which got it from mice. The scientists suggest that OC43 was responsible for a pandemic that killed more than one million people worldwide in 1889–90 – an outbreak previously blamed on influenza. Today, OC43 continues to circulate widely and it might be that continual exposure to the virus keeps the great majority of people immune to it.

## End game

But even if that process made OC43 less deadly, it is not yet clear whether something similar would happen with SARS-CoV-2. A study in monkeys showed that they retained antibodies to SARS-CoV-2, but the researchers only reported on the first 28 days after infection, so it is unclear how long the immunity lasted<sup>12</sup>. Concentrations of antibodies against SARS-CoV also dropped significantly over a two- to three-year period<sup>13</sup>. Whether those lowered levels would be enough to prevent infection or reduce severity has not been tested. Cats, cows, dogs and chickens do not seem to become immune to the sometimes deadly coronaviruses that infect them, leaving veterinarians over the years to scramble for vaccines. Despite all the questions about whether people retain any immunity to SARS-CoV-2, some countries are promoting the idea of giving survivors ‘immunity passports’ to allow them to venture out without fear of being infected or infecting others.

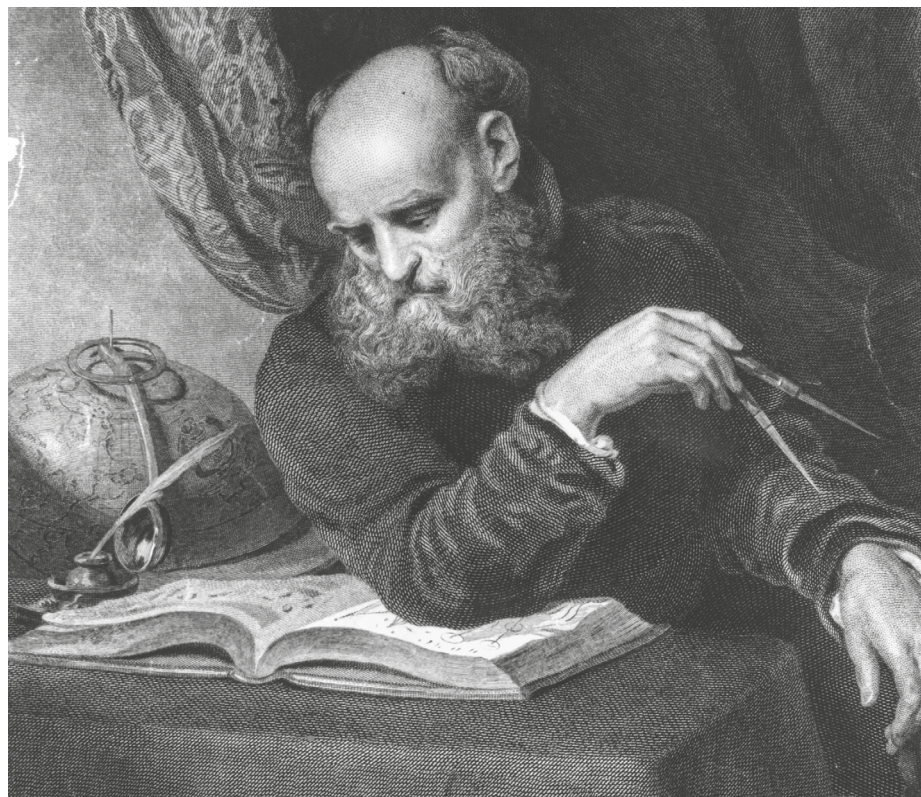
Many scientists are reserving judgement on whether the tamer coronaviruses were once as virulent as SARS-CoV-2. People like to think that “the other coronaviruses were terrible and became mild”, says Perlman. “That’s an optimistic way to think about what’s going on now, but we don’t have evidence.”

David Cyranoski reports for *Nature* from Shanghai, China.

1. Luis, A. D. et al. *Proc. R. Soc. Lond. B* **280**, 20122753 (2013).
2. Brook, C. E. et al. *eLife* **9**, e48401 (2020).
3. Graham, R. L., Donaldson, E. F. & Baric, R. S. *Nature Rev. Microbiol.* **11**, 836–848 (2013).
4. Zhou, P. et al. *Nature* **579**, 270–273 (2020).
5. Zhang, T., Wu, Q. & Zhang, Z. *Curr. Biol.* **30**, 1346–1351 (2020).
6. Boni, M. F. et al. Preprint at bioRxiv <https://doi.org/10.1101/2020.03.30.015008> (2020).
7. Wang, H., Pipes, L. & Nielsen, R. Preprint at bioRxiv <https://doi.org/10.1101/2020.04.20.052019> (2020).
8. Wölfel, R. et al. *Nature* <https://doi.org/10.1038/s41586-020-2196-x> (2020).
9. Wrapp, D. et al. *Science* **367**, 1260–1263 (2020).
10. Callow, K. A., Parry, H. F., Sergeant, M. & Tyrrell, D. A. J. *Epidemiol. Infect.* **105**, 435–446 (1990).
11. Vijgen, L. et al. *J. Virol.* **79**, 1595–1604 (2005).
12. Bao, L. et al. Preprint at bioRxiv <https://doi.org/10.1101/2020.03.13.990226> (2020).
13. Cao, W.-C., Liu, W., Zhang, P.-H. & Richardus, J. H. N. *Engl. J. Med.* **357**, 1162–1163 (2007).



# Books & arts



In the seventeenth century, Galileo faced persecution for his heretical views on astronomy.

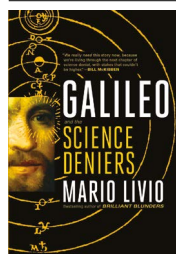
## Galileo's story is always relevant

With science denialism stronger than ever, who better to revisit? **By Alison Abbott**

Is there room in the crowded canon for a new biography of Galileo Galilei? Astrophysicist Mario Livio is betting so. His *Galileo and the Science Deniers* aims to stand out by placing the original Renaissance man and his discoveries in modern scientific and social contexts. In particular, he argues, the charges of heresy that Galileo faced for his scientific claims in the seventeenth century have their counterparts in science deniers' condemnations today.

Born in 1564 in Pisa, Italy, into an intellectual family of declining fortune, Galileo pursued medicine at the University of Pisa. But he soon abandoned his course to study mathematics, his enduring passion. The

Universe, he famously wrote, "is written in the language of mathematics". It was an argot that allowed him to break reliance on the Aristotelian cosmology prized by the Catholic Church, and to forge a new, quantitative study of nature.



**Galileo and the Science Deniers**  
Mario Livio  
Simon & Schuster  
(2020)

In Pisa, Galileo delved into mechanics, using his observations to question accepted ideas about motion (although Livio reminds us that Galileo probably never conducted the famous experiment in which he supposedly dropped spheres from the city's leaning tower and found that they fell at the same speed regardless of their mass).

In 1592, he moved to the University of Padua in Italy, an intellectually liberal environment happily beyond the jurisdiction of the Pope. Here, he began to discuss the revolutionary theory, proposed by Polish mathematician Nicolaus Copernicus in 1543, that rather than being the Universe's fixed centre as Aristotle had insisted, Earth was in fact orbiting the Sun.

Livio structures his account partly around specific works, including Galileo's 1610 *The Sidereal Messenger*, which described his major astronomical observations. While working in Padua, Galileo often visited the nearby port of Venice, where he was introduced to the 'spyglass', a new-fangled instrument from Holland that could be used to see ships approach. Galileo turned it to the heavens to make the discoveries that changed the course of astronomy, and launched his own fate.

In his first observations, Galileo saw that the Moon was not a smooth sphere, but was mountainous. This contradicted the church's view that the heavens were pristine and unchangeable, unlike the corrupt, mutable Earth. He also saw satellites orbiting Jupiter, shooting a hole in the geocentric argument that if Earth were to move, it would lose its Moon.

In 1610, against his friends' advice and in pursuit of more money, Galileo left the protective environment of Padua and moved to Florence to work for Cosimo II de' Medici, the Grand Duke of Tuscany. Despite now living within the Pope's sphere of influence, in 1632, Galileo published his book *Dialogue Concerning the Two Chief World Systems*, an imaginary debate between Salviati, an advocate of heliocentrism, and a witless geocentrist named Simplicio.

Putting the Pope's view – that God's universe is inherently unknowable – into the mouth of a fool was risky. Galileo's fame and fortune rose, but so did the power and determination of his enemies, and the Holy Inquisition finally claimed him. In a riveting account of the trial, Livio describes how the Inquisition dismissed Galileo's claim that the *Dialogue* was a balanced argument that ultimately rejected Copernicanism. On 22 June 1633, one of the world's most venerated scientists was on his knees before its members,

## Books & arts

renouncing the errors and heresies inherent to Copernicanism.

Livio parses the considerable, and often ambiguous, evidence about Galileo's life and trial, and comments on the conclusions reached by various historians. The official summary of the trial proceedings, he writes, "revealed a clear intention to present Galileo in the worst possible light". Like others before him, Livio doubts claims that Galileo left the court defiantly muttering about Earth, "and yet it moves".

The non-chronological zigzagging of the book can be hard to follow, but allows Livio to focus on themes, such as Galileo's polymathy. He highlights Galileo's lifelong study of the great Italian poets Dante Alighieri, Torquato Tasso and Ludovico Ariosto. And he notes that the astronomer's drawing skills and knowledge of perspective allowed him to understand that the shadings on the Moon were shadows cast by mountains, and to depict them in lovely watercolours.

Livio is at his best when he discusses how Galileo's scientific understanding compares

### "Critics from creationists to Donald Trump discredit arguments by exploiting gaps in knowledge."

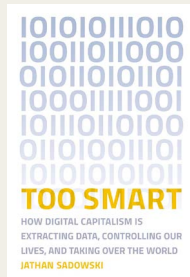
with that of researchers today. Galileo suggested, for example, that comets might be optical phenomena caused by the reflection of sunlight by vapours released from Earth. We now know they are 'dirty snowballs' made of ice, rock, dust and frozen gases. Some of these components vaporize when they get close to the Sun, giving comets two tails: one of dust that reflects sunlight, and one of gas that glows as it ionizes.

And what of today's science deniers? Livio briefly addresses how religion and business interests still conspire to attack evidence for evolution and anthropogenic climate change. In general, "processes that are not fully understood don't constitute flaws", he points out, but critics from creationists to Donald Trump discredit scientific arguments by exploiting gaps in knowledge. It's a chillingly relevant theme, yet the parallels he draws between Galileo's trial and contemporary science wars feel thin, and there's a frustrating lack of examples to demonstrate the continuity of denialism through the centuries.

Nonetheless, Livio has added to the canon an accessible and scientific narrative, in which a profound love for Galileo shines through.

**Alison Abbott** is a writer in Munich, Germany.  
e-mail: [abbott.alison@googlemail.com](mailto:abbott.alison@googlemail.com)

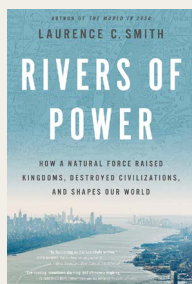
## Books in brief



### Too Smart

Jathan Sadowski MIT Press (2020)

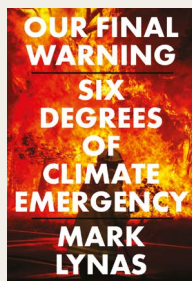
'Smart' technology in hands, homes and cities will "measure, monitor, manage and monetize all aspects of our lives", shows Jathan Sadowski. One toothbrush, for example, uses sensors to record when, how long and how well users brush their teeth, sending the information to cloud servers owned by the manufacturer or a third party. Users might choose to share such data with their dentists — but surely not with dental insurers without express permission. This hype- and jargon-free warning deserves a wide welcome.



### Rivers of Power

Laurence C. Smith Little, Brown Spark (2020)

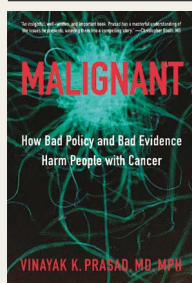
Many civilizations began beside great rivers: the Tigris and Euphrates in what's now Iraq, the Nile in Egypt, the Indus in India and Pakistan, and the Yellow in China. Rivers run through almost all our great cities, notes Earth scientist Laurence Smith in his highly readable history, extolling their fundamental benefits: "access, natural capital, territory, well-being and a means of projecting power". No wonder we still speak of 'crossing the Rubicon', a border river traversed by Julius Caesar's army in defiance of republican law, triggering the rise of the Roman empire.



### Our Final Warning

Mark Lynas Fourth Estate (2020)

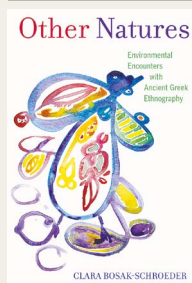
The average global temperature has risen 1°C above pre-industrial levels. This update to environmental activist Mark Lynas's 2007 book *Six Degrees* explores the likely effects of further rises. Its horrifying eloquence derives from its restraint, grounded in research, including a 2019 *Nature* assessment of future carbon dioxide emissions from existing infrastructure, which leads Lynas to forecast total emissions double those required to keep increases to the 1.5°C agreed in Paris in 2015. "This really is our final warning," he concludes.



### Malignant

Vinayak K. Prasad Johns Hopkins Univ. Press (2020)

The cost of new cancer treatments in the United States is routinely US\$100,000 per year per person; some are more than \$400,000 per dose. Oncologist Vinayak Prasad's insider study analyses how US scientific, industrial and regulatory policies — as well as financial conflicts of interest for physicians — "incentivize the pursuit of marginal or unproven therapies at lofty and unsustainable prices". Aimed at general readers (including patients), oncology trainees and experts in health-care policy, it informs and disturbs throughout.



### Other Natures

Clara Bosak-Schroeder Univ. California Press (2020)

US natural-history museums haunt classicist Clara Bosak-Schroeder's study of how ancient Greek historians Herodotus and Diodorus Siculus portrayed non-Greek peoples, such as Ethiopians and Persians. She begins with her childhood fascination with a diorama of a crouching African woman holding a pink grub to her mouth, and ends as she adopts a killer whale at Washington's Whale Museum. Ancient ethnographies, she says, can help people "confront environmental degradation and transform their own relationships to other species". **Andrew Robinson**



# Philip W. Anderson

(1923–2020)

## Nobel winner who transformed condensed-matter and particle physics.

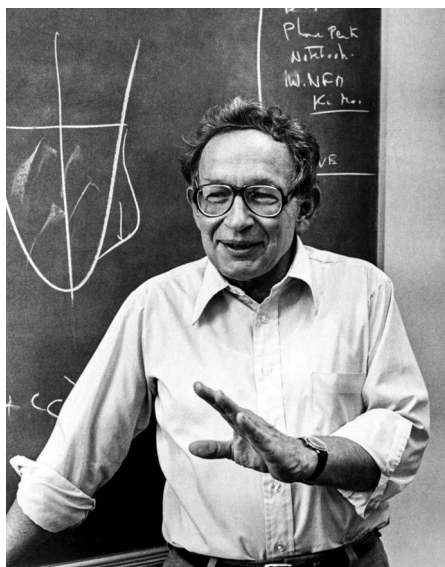
**P**hilip Warren Anderson, who has died aged 96, led the development of condensed-matter physics. In 1977, he won a share of the Nobel Prize in Physics for his discovery of electron localization, whereby disordered metals become insulators, and for his pioneering work on magnetism. His studies of superconductors led him to propose how the force carriers between subatomic particles, such as photons, acquire mass: the Anderson–Higgs mechanism is now part of the standard model of particle physics.

Quantum physics' early triumphs persuaded some that advances would derive exclusively from reducing nature to its most fundamental particles. Anderson rejected this view, arguing that the emergent properties that develop when matter comes together are equally significant. He reasoned that, in science, each new level of complexity requires new fields, connecting physics, chemistry, biology, computer science and economics.

Anderson was born in 1923. The son of a plant pathologist, he grew up in Urbana–Champaign, Illinois, and developed an early passion for speed skating. At 16, he went to Harvard University in Cambridge, Massachusetts, where he fell in love with physics. During the Second World War, he worked on radar at the US Naval Research Laboratory in Washington DC. He first learnt about quantum mechanics when a colleague gave him a precious text on it to repay a wartime loan. Returning to Harvard for graduate work, he studied the effects of pressure on the broadening of spectral lines, mentored by John Hasbrouck Van Vleck, one of those with whom he shared the Nobel prize.

In 1949, William Shockley, co-inventor of the transistor, hired Anderson to join the theory group at Bell Telephone Laboratories in Murray Hill, New Jersey. Here, Anderson first focused on magnetism. Quantum mechanics predicts that electrons carry tiny magnetic moments called spin. The attraction of fridge magnets results when spins orient in the same direction. By contrast, the read heads of hard disk drives rely on an antiferromagnet, in which the spin alternates direction on adjacent atomic layers. By combining the effects of electron repulsion and quantum mechanics, Anderson explained how iron atoms become magnetic, and accounted for the interactions that result in antiferromagnetism.

In 1958, Anderson discovered a remarkable parallel between magnetism and



superconductivity. Whereas magnets concentrate fields, superconductors expel them, allowing them to levitate. Following John Bardeen, Leon Cooper and Robert Schrieffer's discovery that superconductivity results from electrons forming Cooper pairs, Anderson realized that these pairs are a kind of pseudospin. In a magnet, fluctuations in the magnetization propagate, forming a spin wave. Cooper pairs are charged, and pseudospin waves cause an electric current that interacts with the electromagnetic fields, which are carried by photons. When Anderson calculated the motion of photons inside a superconductor, he found that they acquired a mass.

Anderson realized that there were still deeper parallels between superconductivity and particle physics. In 1962, he proposed a mechanism for subatomic force carriers called gauge bosons to acquire mass in a kind of cosmic superconductor, now known as the Higgs field after the British physicist Peter Higgs. Anderson's work was prominently cited in Higgs's 1964 paper predicting the existence of the Higgs boson.

Work on semiconductors at Bell Labs led Anderson to propose in 1957 that disorder – such as that caused by defects and impurities in a material – localizes electron waves, producing an insulator. Today, Anderson localization is recognized as a general property of all kinds of waves in disordered media. But the original idea was radical and took two decades and the contributions of many leading physicists to be developed in detail. Anderson also studied spin

glasses, a type of random magnet that contains both ferromagnetic and antiferromagnetic interactions. Working with Sam Edwards at the University of Cambridge, UK, he modelled the glass's capacity to 'remember' the history of its environment. The Edwards–Anderson model was an early forerunner of the neural networks used in modern machine learning.

Anderson had an artist's eye for original interpretation, and worked closely with experimentalists to develop ideas that led to new fields of study. During a year in Japan in 1953, visiting the mathematical physicist Ryogo Kubo at the University of Tokyo, he became a master in the ancient game of Go. Just as he could instinctively see several moves beyond his opponent, he often reached an understanding of physics that was hard for others to grasp. Unusually intuitive for a theoretical physicist, he was able to reduce complex problems so they would succumb to a minimum of mathematics.

To the many younger physicists he mentored, with whom he generously shared inspiring ideas, he was something of a guru. He would invite them, and their families, for dinner, developing lasting friendships. He had a passion for the outdoors, spending Sundays gardening and clearing the woodland around his house.

From 1967 to 1975, Anderson was a visiting professor at Cambridge, where he was a close associate of Nevill Mott, the third of the 1977 Nobel prizewinners. In 1975, he took up a position at Princeton University in New Jersey, where he devoted much of his research to high-temperature superconductivity (discovered in 1986 by Georg Bednorz and Alex Müller).

Anderson's resonating valence bond (RVB) theory, proposed in 1987, once again borrows ideas from magnetism, positing that high-temperature superconductivity results from the injection of charge into an insulating quantum spin liquid. High-temperature superconductivity has still not been achieved, but many think that RVB theory contains the seeds of how it might be. Although Anderson has left us, his ideas are still many moves ahead.

**Piers Coleman** is a distinguished professor at the Rutgers Center for Materials Theory, Rutgers, the State University of New Jersey. He was a graduate student with Philip W. Anderson at Princeton in 1980–1984. e-mail: coleman@physics.rutgers.edu

# Correspondence

## COVID-19: remote tech spares clinicians

Virtual-reality technology backed by an ultrafast 5G mobile network (VR+5G) has complemented rigorous preventive measures in cutting the infection rate of medical staff on the front line of the COVID-19 epidemic at a university hospital in Zhejiang, China. This technology enables remote diagnosis and treatment monitoring of patients, thereby helping to reduce direct contact with medical staff (see [go.nature.com/35ex2pv](https://go.nature.com/35ex2pv)).

This VR system allows medical personnel to interact with a digital 3D representation of the patient's environment through a computer-generated simulation in real time. Tracking data collected through such VR+5G applications could also provide valuable information for policymakers, for example on factors that influence the course, infectivity and severity of the illness.

VR+5G would also improve the monitoring and communication capability of people at home.

**Guoen Cai** Fujian Medical University Union Hospital, China.

**Guofa Cai** Guangdong University of Technology, China.  
[caiguofa2006@gdut.edu.cn](mailto:caiguofa2006@gdut.edu.cn)

## Arab students thrive in Israeli university

For long-standing reasons, the Arab minority in Israel has been under-represented in Israeli universities. The Technion — Israel Institute of Technology has rectified this by developing an empowerment model that is now being extended to all other academic institutes in Israel through the Israeli Council for Higher Education.

The Technion resolved two weaknesses, without resorting to US-style affirmative action to resolve discrimination. These were the inadequacy of Arab students' secondary-school preparation for higher education, and the difficulty of adjusting to campus life in the face of cultural and age-related differences from the Jewish majority. The Technion set up coaching programmes for prospective students, workshops in core courses run by high-achieving Arabic-speaking students, personal tutoring in social engagement, and professional guidance on self-management.

Since 2004, our numbers of Arabic-speaking students and Arab female students have increased by 200% and 350%, respectively. The number of Arab students deemed 'outstanding' has risen by 1,800%, and the number of Arab graduate students by 120%. Dropout rates among our Arab students have fallen by 67%.

**Hossam Haick, Peretz Lavie**  
Technion — Israel Institute of Technology, Haifa, Israel.  
[hossam@technion.ac.il](mailto:hossam@technion.ac.il)

**Fatima Abreek-Zubiedat** Swiss Federal Institute of Technology (ETH), Zurich, Switzerland.

## Learn truths from replicability testing

More incentive is needed to spur investigation into replication issues and null results (see *Nature* **578**, 489–490; 2020). For example, experienced scientists could encourage junior researchers to allocate part of their time to verifying other researchers' results, which would also provide them with essential insights into the scientific method. To support such ventures, we are launching *ReScienceX*, a free-to-publish and free-to-read, peer-reviewed journal that will be devoted to reproduction and replication experiments.

Research students and early-career researchers should be given credit for publishing replication studies. This would have the advantage of also providing them with valuable research training in experimental design and the use of analytical tools. It is currently not easy for researchers aspiring to tenured positions to divert from generating high-impact publications, which are still the prime measure of performance (*Nature Hum. Behav.* **3**, 1001; 2019). In our view, however, they could significantly contribute to scientific endeavour by incorporating rigorous replication studies into their daily routine and embracing null results.

**Etienne B. Roesch** University of Reading, Reading, UK.  
[e.b.roesch@reading.ac.uk](mailto:e.b.roesch@reading.ac.uk)

**Nicolas Rougier** Inria Bordeaux South West Research Centre, Talence, France.

## Don't underrate high-flyer SOFIA

I find your assessment of the scientific worth of the Stratospheric Observatory for Infrared Astronomy (SOFIA) overly harsh (*Nature* **580**, 314; 2020). The facility provides invaluable data that would otherwise be unobtainable until well into the future.

In addition to the unique features you mention, SOFIA is the only observatory open to the community that operates in the infrared, from 28 to 300 micrometres. Data from these wavelengths harbour clues about star formation from interstellar reservoirs of cold gas and dust, hinting at how our Solar System came into being. They also hold information about stellar death, the entourage of massive black holes in galactic nuclei and interstellar chemistry.

Balloon platforms can operate in this spectral region, but do not offer SOFIA's diverse observing opportunities. The observatory's instruments are much in demand, particularly the far-infrared camera and polarimeter, and the GREAT heterodyne spectrometer. These generate 40 or so papers annually from this still-young facility.

There are plans to launch satellite observatories that could replace SOFIA — including a joint Japanese–European project called the Space Infrared Telescope for Cosmology and Astrophysics, or SPICA, and the US Origins Space Telescope. However, the cost of these would far exceed the cost of continuing to maintain SOFIA. Furthermore, neither of these observatories could be launched until well into the 2030s.

**Mark R. Morris** University of California, Los Angeles, USA.  
[morris@astro.ucla.edu](mailto:morris@astro.ucla.edu)



# News & views

## Neurodegeneration

# Lipid carrier breaks barrier in Alzheimer's disease

Makoto Ishii & Costantino Iadecola

People who carry the gene variant *APOE4* are at higher-than-average risk of developing Alzheimer's disease. It emerges that this variant is linked to defects in the blood–brain barrier and subsequent cognitive decline. **See p.71**

The best-known hallmarks of Alzheimer's disease are clumps of misfolded amyloid- $\beta$  (A $\beta$ ) and tau proteins, which aggregate in the brain. However, there is increasing awareness that A $\beta$  and tau might not be the whole story – alterations in the blood–brain barrier (BBB) have also emerged as early markers of this neurodegenerative disorder<sup>1</sup>. The degree of disruption to the BBB correlates with the degree of cognitive dysfunction that a person experiences<sup>2</sup>, but what causes BBB breakdown has been unknown. Montagne *et al.*<sup>3</sup> present evidence on page 71 that the leading genetic risk factor for Alzheimer's disease, apolipoprotein E4, is linked to BBB breakdown.

The gene *apolipoprotein E* (*APOE*) encodes a major lipid-carrier protein, ApoE, in the brain<sup>4</sup>. There are three predominant variants of *APOE*: *APOE2*, *APOE3* and *APOE4*. As with almost all genes, people carry two copies of *APOE*, which can be either the same or different variants. Compared with the more-common *APOE3* variant, *APOE4* markedly increases the risk of Alzheimer's disease – up to 4-fold in people with one copy of this variant, and 15-fold in people who have two copies<sup>4</sup>. People carrying *APOE4* who do contract Alzheimer's disease also tend to develop symptoms of the disorder earlier than those who develop the disease but do not carry the variant<sup>4</sup>.

Proteins from blood plasma have been found in the cerebrospinal fluid (the liquid that surrounds the brain and spinal cord) of cognitively healthy people who carry *APOE4* and who subsequently go on to develop Alzheimer's disease. These proteins have presumably leaked through the BBB, indicating that the integrity of the barrier is lost before cognition declines<sup>5</sup>. Evidence from mouse models, and from the brains of people who have died with Alzheimer's disease,

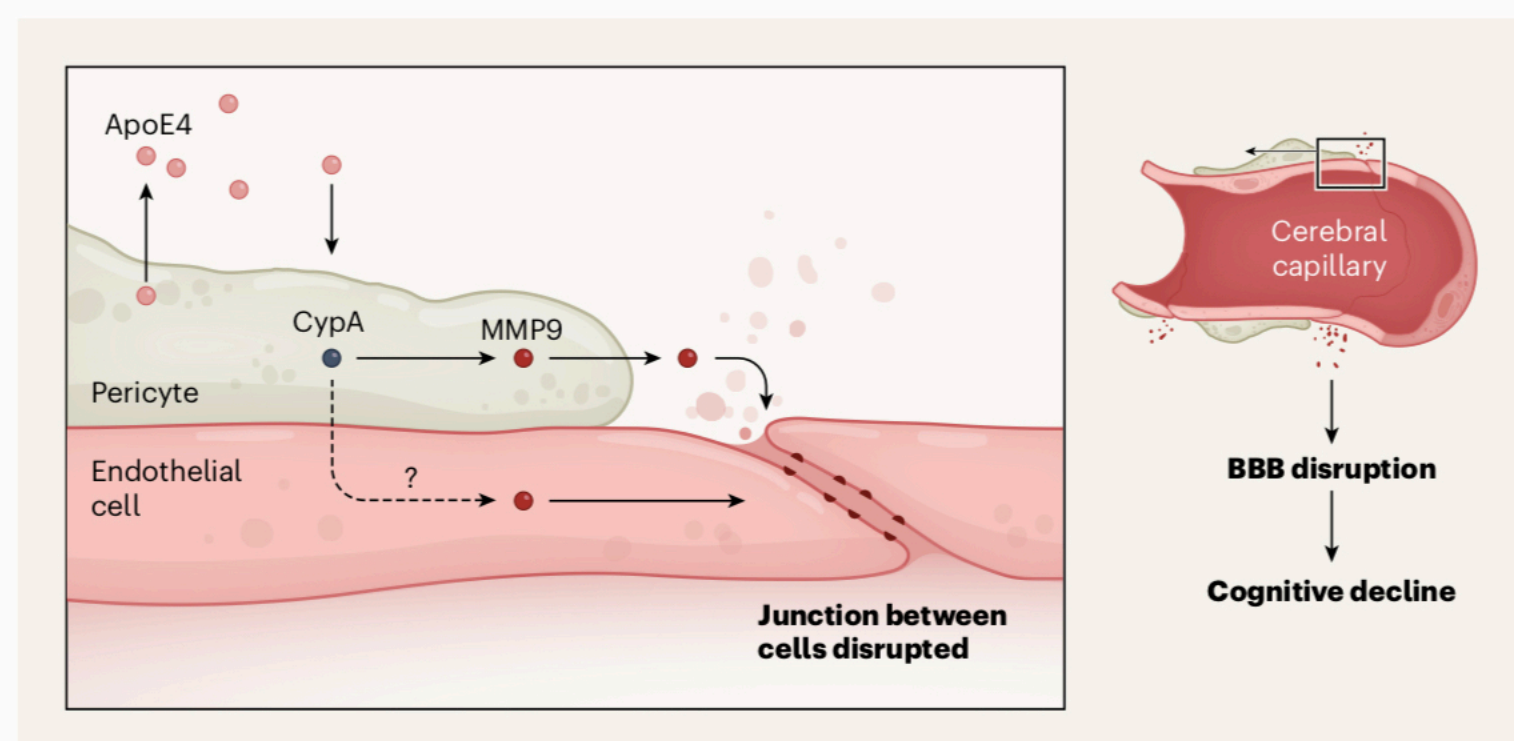
suggests that BBB breakdown is caused by the degeneration of pericytes – cells nestled in the wall of cerebral capillaries. These cells normally safeguard the BBB<sup>5</sup> by preventing the breakdown of junctions between endothelial cells, which make up the capillary walls.

Whether ApoE4 is responsible for early BBB dysfunction in Alzheimer's disease, by itself or in concert with A $\beta$  and tau, was unknown. Montagne and colleagues set out to address this knowledge gap. The authors used a technique called dynamic contrast-enhanced magnetic resonance imaging to investigate the permeability of the BBB in people who had either healthy cognition or mild cognitive impairment (a prelude to Alzheimer's

disease), grouped according to their *APOE* status. They found that people who were cognitively healthy and carried either one or two copies of *APOE4* had a leaky BBB in two brain regions important for memory and cognition – the hippocampus and the parahippocampal gyrus. This leakage was worse in *APOE4* carriers who exhibited mild cognitive decline.

Remarkably, these effects preceded any signs of tissue loss in the hippocampus and parahippocampal gyrus, attesting to the idea that BBB disruption is an early event in the onset of neurodegeneration. BBB leakage was independent of A $\beta$  and tau accumulation, which the authors assessed both by studying samples of cerebrospinal fluid and through another brain-imaging technique, positron emission tomography. Montagne and co-workers found that, unlike in *APOE4* carriers, the BBB was intact in cognitively healthy *APOE3* carriers. It was, however, leaky in *APOE3* carriers who showed cognitive impairment – although less so than in *APOE4* carriers at an equivalent stage of impairment.

Next, Montagne *et al.* examined whether BBB breakdown in *APOE4* carriers was linked to pericyte degeneration. In support of this idea, they found that a biomarker of pericyte injury – a soluble form of a protein known as platelet-derived growth factor-receptor- $\beta$  (sPDGFR $\beta$ ) – was elevated in the cerebrospinal fluid of *APOE4* carriers compared with *APOE3* carriers. High levels of the protein in people who carried *APOE4* were associated with a



**Figure 1 | The gene variant *APOE4* and Alzheimer's disease.** People who carry *APOE4* are at heightened risk of Alzheimer's disease. Montagne *et al.*<sup>3</sup> provide evidence that ApoE4 protein is secreted by cells called pericytes, which abut endothelial cells that line cerebral capillaries at the blood–brain barrier (BBB). Secreted ApoE4 activates the protein cyclophilin A (CypA) in the pericytes. This triggers a downstream signalling pathway involving activation of the inflammatory protein matrix metalloproteinase-9 (MMP9) in pericytes, and possibly also in endothelial cells. This causes disruption of junctions between adjoining endothelial cells, opening the BBB in brain regions involved in learning and memory. Disruption of the BBB is associated with impaired cognition, although the mechanisms that link the two are unclear.



leaky BBB and cognitive impairment. sPDGFR $\beta$  elevation was independent of A $\beta$  and tau.

The authors then looked for insight into the mechanisms by which pericytes might become injured. They focused on cyclophilin A (CypA) and matrix metalloproteinase-9 (MMP9), two proteins that are part of an inflammatory pathway implicated in *APOE4*-driven pericyte damage and BBB breakdown<sup>6</sup>. Levels of CypA and MMP9 in the cerebrospinal fluid were higher in *APOE4* carriers who had mild cognitive impairment than in cognitively healthy *APOE4* carriers or *APOE3* carriers who had comparable cognitive dysfunction. Again, this change was not related to increases in A $\beta$  or tau.

Finally, the researchers generated pericytes *in vitro* from human induced pluripotent stem cells that expressed *APOE3* or *APOE4*. They found that *APOE4*-expressing pericytes secreted substantially more CypA and MMP9 than did *APOE3* pericytes. ApoE4 (but not ApoE3) secreted by pericytes activates the CypA–MMP9 pathway on nearby pericytes – the cells therefore cause their own demise. ApoE4 could also activate the CypA–MMP9 pathway in endothelial cells, which are susceptible to the harmful effects of *APOE4* (ref. 7). Therefore, injury to pericytes and endothelial cells might both cause BBB leakage (Fig. 1).

These observations cast new light on *APOE4* that runs contrary to the widely held idea that this gene variant contributes to Alzheimer's disease solely by promoting A $\beta$  and tau accumulation<sup>4</sup>. Instead, it seems that BBB dysfunction might explain why *APOE4* carriers are susceptible to Alzheimer's disease. The authors' findings might also explain why *APOE4* carriers have worse outcomes following stroke or traumatic brain injury<sup>8</sup> than do people who carry other *APOE* variants. However, as Alzheimer's disease progresses, *APOE4* could also slow A $\beta$  and tau clearance, exacerbating declines in cognition.

Even more striking is the finding that early drivers of cognitive impairment differ between *APOE4* and *APOE3* carriers. Montagne and colleagues' findings indicate that activation of the CypA pathway and pericyte damage might not be involved in cognitive impairment in people who carry the most common *APOE* variant, *APOE3*. But whether a leaky BBB caused by factors that are independent of pericytes (for example, damage to endothelial cells caused by A $\beta$ ; ref. 1) contributes to cognitive impairment in *APOE3* carriers remains unclear. The role of the BBB in *APOE2* carriers, which was not assessed in the current study, also remains unknown. Although *APOE2* is associated with a reduced risk of Alzheimer's disease compared with other *APOE* variants, this is unlikely to result from a more resilient BBB, because *APOE2* carriers have an increased risk of microhaemorrhages, suggesting vascular frailty<sup>4</sup>.

Whether and how BBB breakdown leads to cognitive impairment also remains to be

determined. Is it a cause or a consequence of the disease process? Evidence from mice indicates that some proteins in the blood, such as fibrinogen, damage the synaptic connections between neurons<sup>9</sup>. But a pathogenic role for these proteins in the human brain has not yet been demonstrated.

Irrespective of these questions, Montagne *et al.* have broadened our understanding of how *APOE4* promotes cognitive impairment. They have also demonstrated that different *APOE* statuses can promote disease through different mechanisms. A deeper appreciation of how gene variants shape Alzheimer's disease might prove crucial for more-personalized approaches to treating this prevalent and incurable disease.

## Atomic physics

# Exotic helium atom lit up

Niels Madsen

An elusive type of atom known as pionic helium has been directly excited by laser light for the first time. The work establishes a promising experimental platform for probing fundamental physics. **See p.37**

Exotic atoms are those in which one or more of the constituents of normal atoms have been replaced by an exotic particle, such as an antimatter particle. These atoms can then be probed to search for any tiny discrepancies in their properties from those predicted by models using techniques that underpin the world's most accurate timekeepers, atomic clocks – and thereby opening a window on the foundations of physics. On page 37, Hori *et al.*<sup>1</sup> are the first to report laser excitation of helium atoms in which one electron has been replaced by a subatomic particle called a pion.

The interest in exotic atoms arises from the fact that they often facilitate the most basic experimental strategy used in physics: changing a single parameter or component in an otherwise complex system, to observe the effect. In practice, this is not as simple as it might seem. Different particles can have different masses or charges, and might interact with their surroundings differently in other subtle ways. However, such subtleties often add to the value of exotic atoms.

As the techniques needed to study exotic atoms improve, increasing numbers of scientists are working with these atoms to investigate the fundamental properties of nature. A good example of this is the 'proton radius puzzle', which arose from a study of muonic hydrogen<sup>2</sup> – a hydrogen atom in which the electron has been replaced by a subatomic

**Makoto Ishii** and **Costantino Iadecola** are at the Feil Family Brain and Mind Research Institute, Weill Cornell Medicine, New York, New York 10065, USA.  
e-mail: coi2001@med.cornell.edu

1. Cortes-Canteli, M. & Iadecola, C. *J. Am. Coll. Cardiol.* **75**, 942–951 (2020).
2. Nation, D. A. *et al. Nature Med.* **25**, 270–276 (2019).
3. Montagne, A. *et al. Nature* **581**, 71–76 (2020).
4. Yamazaki, Y., Zhao, N., Caulfield, T. R., Liu, C.-C. & Bu, G. *Nature Rev. Neurol.* **15**, 501–518 (2019).
5. Profaci, C. P., Munji, R. N., Pulido, R. S. & Daneman, R. *J. Exp. Med.* **217**, e20190062 (2020).
6. Bell, R. D. *et al. Nature* **485**, 512–516 (2012).
7. Koizumi, K. *et al. Nature Commun.* **9**, 3816 (2018).
8. Mahley, R. W., Weisgraber, K. H. & Huang, Y. *Proc. Natl Acad. Sci. USA* **103**, 5644–5651 (2006).
9. Merlini, M. *et al. Neuron* **101**, 1099–1108 (2019).

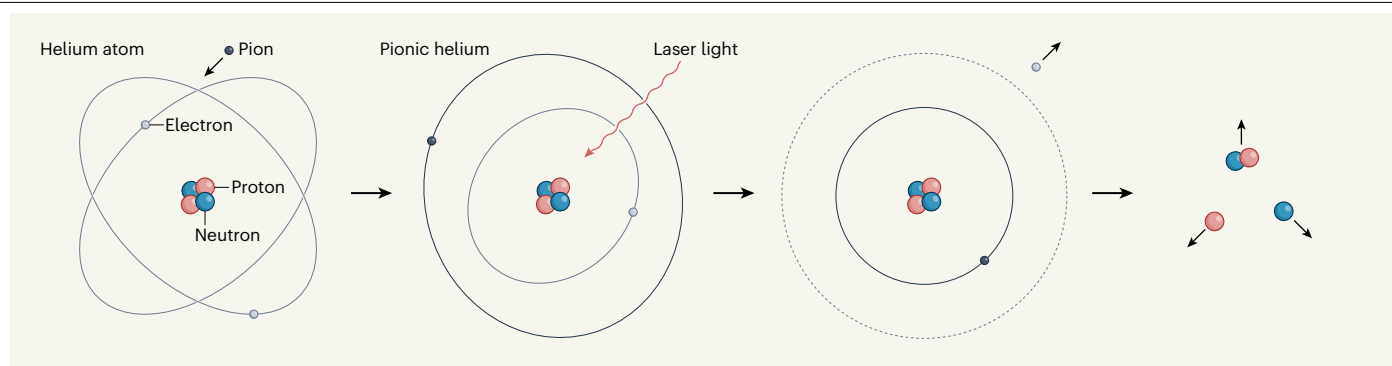
This article was published online on 29 April 2020.

muon particle (muons have similar properties to electrons, but have about 200 times greater mass).

Muonic hydrogen was used to determine a key property of the proton known as the charge radius, but the value obtained was about seven standard deviations away from the expected value at the time. The value obtained using muonic hydrogen has since been independently confirmed in a study of ordinary hydrogen<sup>3</sup>, and also in experiments in which electrons are scattered from protons<sup>4</sup>, potentially clarifying the true value of the proton radius and thus solving the puzzle. Nevertheless, muonic hydrogen aptly illustrates how exotic, sometimes short-lived, atomic systems can be used to poke holes in seemingly well-established results.

An important feature of exotic atoms that adds to their utility as probes for fundamental physics is that they are bound systems (energy is needed to pull their components apart), with multiple internal energy states. Transitions between these states are therefore amenable to study by laser spectroscopy, the most precise measurement tool in the physics toolkit. The study of transitions in atoms – and particularly in the hydrogen atom – is an ongoing effort that has spanned more than two centuries. It inspired Niels Bohr's groundbreaking model of the atom in the early twentieth century, for example, and has driven much





**Figure 1 | How to make and excite pionic helium.** Hori *et al.*<sup>1</sup> fired a beam of negatively charged subatomic particles, called pions, at liquid helium. When a pion struck a helium atom, it could knock out one of the electrons and replace it in a high-energy orbit around the helium nucleus (which consists of two protons and two neutrons), thus forming an exotic atom known as pionic helium. The authors fired laser light at the exotic atoms and thus observed the transition of the pion to a lower-energy orbit – a process that triggers the ejection of the remaining electron. This ejection speeds up the absorption of the pion by the nucleus (not shown), which finally breaks apart.

of the development of quantum mechanics.

Today, atomic transitions are the foundation on which all time measurements are built: a transition in the caesium-133 atom provides a reference value that underpins the International System of Units definition of the second. Therefore, spectroscopic techniques for measuring transitions are constantly being refined, and the best measurements can now reach staggering precisions – to almost 18 significant figures<sup>5</sup>. The most precise measurements of exotic atoms still lag some way behind, but a measurement on antihydrogen (the bound state of an antiproton and an anti-electron) achieved a milestone precision of almost 12 significant figures<sup>6</sup>, thus paving the way for extremely sensitive tests of the fundamental properties of antimatter.

In their seminal work, Hori *et al.* record the first observation of a transition in a ‘pionic’ helium atom. In such an atom, one of the two electrons of a helium atom has been replaced by a subatomic particle called a pion. Pions were discovered<sup>7</sup> by Cecil Powell and co-workers in 1947, but their existence was first predicted<sup>8</sup> in 1935 by Hideki Yukawa. They belong to the family of subatomic particles known as mesons, which are made up of a quark and an antiquark; quarks are the particles that make up protons and neutrons.

Pions are short-lived particles that come in positively charged, negatively charged and neutral varieties. The negatively charged pions used by Hori *et al.* have a lifetime of only 26 nanoseconds when isolated. It is thus no small feat that the experimenters not only succeeded in replacing an electron in helium atoms with a pion, but also observed the resulting exotic atom undergo a quantum transition. A further difficulty is that the lifetime of the pion in the exotic atom can be reduced to picoseconds because of its vicinity to the atom’s nucleus.

The authors prepared pionic helium atoms by firing a beam of pions at a liquid-helium target. In Hori and colleagues’ experiment,

the helium target was cooled to a cryogenic temperature of about 2 kelvin. This allowed some pions to be captured in a weakly bound state of pionic helium, in which the pion was sufficiently far from the nucleus to be shielded from it by the remaining electron (Fig. 1). The resulting exotic atom therefore retained a lifetime of nanoseconds, which is long enough for a laser pulse to excite the nascent exotic atom.

Using a short pulse (0.8 nanoseconds) of infrared laser light, Hori and colleagues provoked a transition of the pion. This resulted in the ejection of the remaining electron, leaving a short-lived system consisting of just a pion bound to a helium nucleus. The pion was then

**“The authors prepared pionic helium atoms by firing a beam of pion particles at a liquid-helium target.”**

absorbed by the nucleus, leading to the latter’s break-up (fission).

The pion transition was detected by carefully removing large ‘background’ signals from the experimental data; this background was associated with fission products from short-lived states of pionic helium, or was generated by the pion beam itself. That left a signal from just three transitioning pionic helium atoms per hour, or an estimated three per billion exotic atoms produced. Despite the challengingly low numbers, the laser-induced signal for the transition was clearly detected, and the laser frequency at which it occurred (and which corresponds to the energy change of the transition) could be determined with an absolute precision of about five significant figures.

This result builds on Hori and colleagues’ extensive experience of studying another exotic helium-type atom, in which one electron was replaced by an antiproton<sup>9</sup>. That work resulted in, among other things, the most

precise determination so far of the ratio of the mass of the antiproton to that of the electron. However, the authors had to overcome further challenges to study pionic helium.

For example, the lifetimes of pionic helium atoms are shorter than those of antiprotonic helium and the widths of lines in its spectrum are broadened, because the excited exotic atoms undergo many collisions in the relatively dense liquid-helium target. It might be possible to reduce the density of the liquid helium to resolve these somewhat incalculable effects, although this would also lower the signal rate. It remains to be seen whether extrapolation to zero density, where collisional effects are minimized, is practicable.

Nevertheless, Hori and colleagues’ work opens up a whole new experimental system for further exploration. If the above challenges can be overcome, such exploration might enable the accuracy of the mass of the negative pion to be improved by a factor of 10–100, for instance; currently, this mass is known to only six decimal places. The experiment thus paves the way to fresh insights into the fundamental constituents of nature.

**Niels Madsen** is in the Department of Physics, Swansea University, Swansea SA2 8PP, UK. e-mail: n.madsen@swansea.ac.uk

1. Hori, M., Aghai-Khozani, H., Sôtér, A., Dax, A. & Barna, D. *Nature* **581**, 37–41 (2020).
2. Pohl, R. *et al.* *Nature* **466**, 213–216 (2010).
3. Bezginov, N. *et al.* *Science* **365**, 1007–1012 (2019).
4. Xiong, W. *et al.* *Nature* **575**, 147–150 (2019).
5. McGrew, W. F. *et al.* *Nature* **564**, 87–90 (2018).
6. Ahmadi, M. *et al.* *Nature* **557**, 71–75 (2018).
7. Lattes, C. M. G., Muirhead, H., Occhialini, G. P. S. & Powell, C. F. *Nature* **159**, 694–697 (1947).
8. Yukawa, H. *Proc. Phys. Math. Soc. Japan* **17**, 48–57 (1935).
9. Hori, M. *et al.* *Science* **354**, 610–614 (2016).



# Hidden in plain sight

Anirban Maitra

Pancreatic cancer does not respond to certain anticancer treatments that boost immune responses. A mechanism active in tumour cells that contributes to this evasion of immune targeting has been uncovered. **See p.100**

A consistent hallmark of pancreatic cancer is the inability to treat it with immunotherapy – an approach that harnesses the body's immune response to target a cancer. On page 100, Yamamoto *et al.*<sup>1</sup> reveal a mechanism that enables pancreatic cancer cells to evade an immune response. The process implicated is usually associated with the normal degradation and recycling of cellular proteins. The authors find that inhibition of this pathway, using drugs or by genetic approaches, reverses this immune evasion in animal models of pancreatic cancer. The finding provides a compelling rationale for investigating whether targeting this pathway might be of benefit in the clinic.

The past decade has seen striking advances in the use of immunotherapy to treat numerous solid cancers (those not formed from blood cells) that had not responded to earlier therapy attempts. One such advance is the development of what is known as checkpoint blockade therapy. This targets the proteins PD-1 and CTLA-4, which are found on the surface of immune cells called cytotoxic T cells (also called CD8 T cells) and inhibit an immune response. The administration of two antibodies that, respectively, target these two inhibitory proteins is used widely to drive an antitumour immune response<sup>2</sup>. Unfortunately, treatment of pancreatic cancer with either antibody, or the two together, has not led to any notable success in terms of patient survival<sup>3</sup>.

There are many factors that can contribute to the failure of immunotherapy. The tumour microenvironment of pancreatic cancer contains a variety of immune cells that suppress the function of cytotoxic T cells – they include myeloid cells, tumour-associated macrophages and regulatory B cells<sup>4</sup>. Moreover, signals released by the cancer cells themselves (often mediated by the aberrant activation of tumour-promoting genes such as those encoding the proteins Ras and Myc) have a key role in creating this profoundly immunosuppressive environment<sup>5</sup>. In addition, fibroblast cells in the tumour microenvironment secrete material that generates a

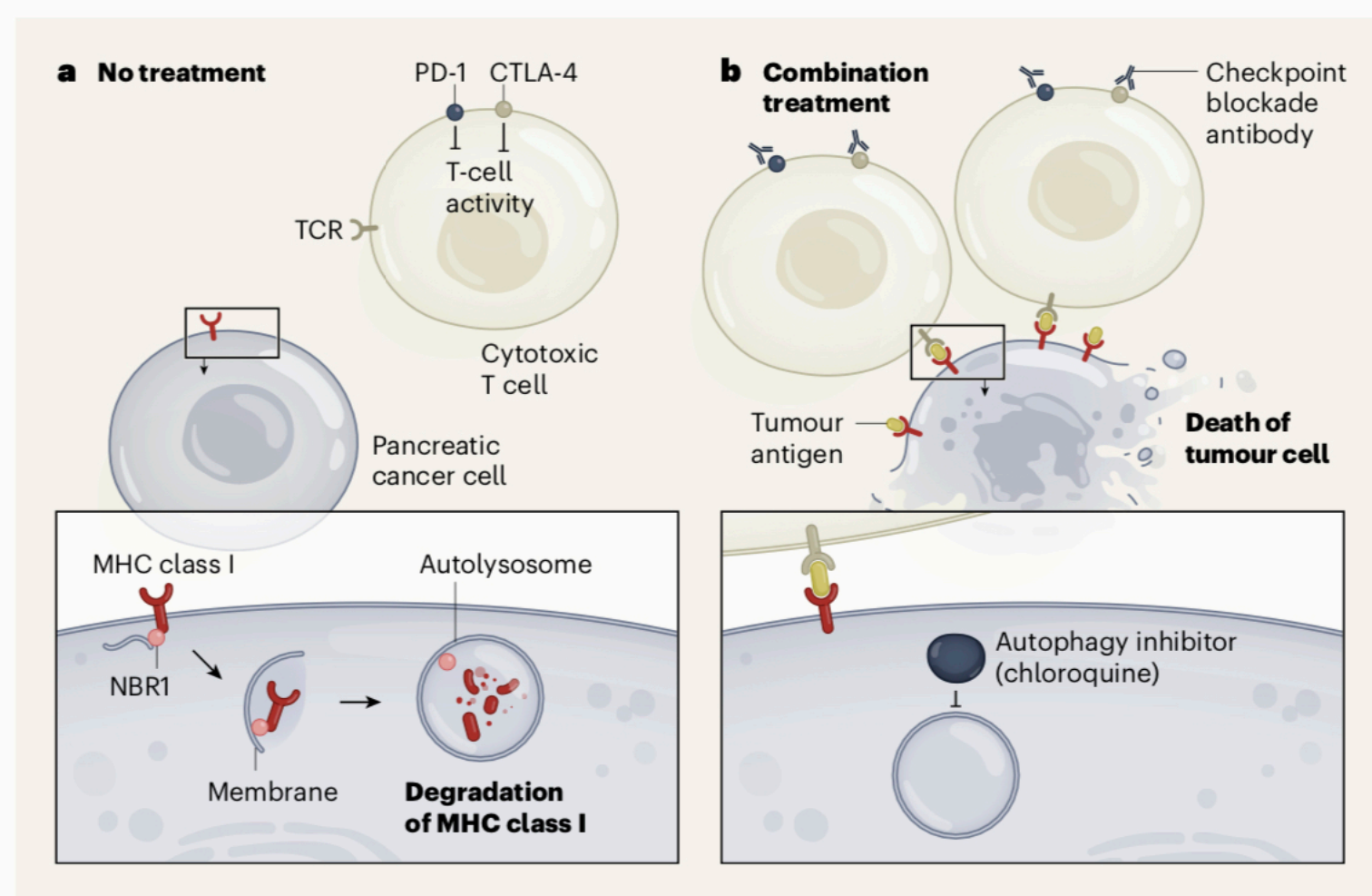
physical barrier hindering the influx of T cells to the tumour site<sup>6</sup>.

Another equally crucial cause of immune evasion is that the cancer cells themselves undergo changes. An immune response is triggered when a T cell recognizes as foreign a peptide fragment (called an antigen) that is 'presented' on the surface of a cancer cell, bound to a molecule called a major histocompatibility complex (MHC) class I molecule (Fig. 1). An inability of the MHC molecules to present tumour antigens has emerged<sup>7</sup> as an explanation for how cancer cells can hide from the immune system. These molecules

comprise an invariant  $\beta_2$ -microglobulin protein (encoded by the *B2M* gene) and a protein encoded by HLA genes, which vary in the antigen-binding site they encode, thereby enabling different antigens to be bound by different MHC molecules. Tumour-derived-antigen presentation by MHC class I molecules is required for cancer cells to be recognized by the T-cell receptor on cytotoxic T cells.

The inability to present antigens, and thus the absence of immune recognition, can occur through alterations in the genes encoding proteins needed for antigen presentation, such as *B2M* and HLA genes<sup>8,9</sup>. In pancreatic cancer, such genetic alterations are relatively uncommon, occurring in no more than 1% of cases<sup>10</sup>. However, lower-than-normal levels of MHC class I molecules, or their complete loss, occurs in more than 60% of cancers that arise in the pancreas<sup>11</sup>, and the decrease might be even greater in metastatic tumours (those that have spread beyond the pancreas)<sup>7</sup>. Until now, the mechanisms underlying the regulation of MHC class I molecules in pancreatic cancer have remained elusive.

Yamamoto and colleagues reveal that a cellular pathway called autophagy is the means by which pancreatic cancer cells limit the amount of MHC class I molecules on their surface, thereby hindering antigen



**Figure 1 | Boosting the targeting of pancreatic cancer by the immune system.** The ability of immune cells called cytotoxic T cells to attack tumours is thwarted by, for example, the action of inhibitory proteins called CTLA-4 and PD-1. Yamamoto *et al.*<sup>1</sup> report a previously unsuspected mechanism that enables pancreatic cancer to evade immune cells. **a**, The T-cell receptor (TCR) on the surface of cytotoxic T cells enables them to recognize tumour cells, and this recognition depends on the presence of major histocompatibility complex (MHC) class I molecules on the surface of cancer cells. Yamamoto and colleagues report that MHC class I molecules in pancreatic cancer cells are destroyed by a process termed autophagy. This begins when the molecule binds to the protein NBR1 and is enveloped in a membrane to form an organelle called an autolysosome. The destruction of MHC class I molecules thus prevents them from 'presenting' peptide fragments (antigens) from the tumour that might be recognized by the TCR. **b**, In mouse models of pancreatic cancer, treatment combining an autophagy-inhibiting drug (chloroquine) and antibodies targeting PD-1 and CTLA-4 (termed checkpoint blockade therapy) provoked a robust immune response against the tumour compared, with the case for animals that did not receive such treatment.



presentation. Autophagy is an essential cellular degradation pathway that recycles organelles and proteins to maintain cellular ‘fitness’<sup>12</sup>. It can act selectively through the binding of a specific receptor to a ‘cargo’ (for example, a protein or organelle) that has been targeted for destruction by being marked with a tag, such as the protein ubiquitin. A complex of receptor and cargo is enveloped in a lipid membrane to form a vesicle called an autophagosome, which fuses with an organelle known as a lysosome to form an autolysosome (Fig. 1). The cargo then undergoes enzyme-mediated digestion in the autolysosome and its contents are recycled for use in the cell.

Yamamoto *et al.* report that, remarkably, in pancreatic cancer cells, most MHC class I molecules do not exist on the cell surface, but instead are found in autophagosomes and autolysosomes. The authors identified an autophagy-associated receptor called NBRI as being responsible for targeting MHC class I molecules to the autophagy machinery. Furthermore, they found that if autophagy was inhibited in mice, by drugs such as chloroquine or through genetic engineering, this restored the surface expression of MHC class I molecules, thereby enhancing antigen presentation. In mouse models of pancreatic cancer, autophagy inhibition resulted in an influx of cytotoxic T cells to the tumour microenvironment, and if the animals also received checkpoint blockade therapy, a robust antitumour immune response was generated.

Increased autophagy has been known for around a decade<sup>13</sup> to be a metabolic requirement for pancreatic cancer, but only now has a connection been made to the immune evasion of tumour cells. Thus far, clinical trials targeting autophagy in pancreatic cancer have relied on testing the antimalarial drug hydroxychloroquine (chloroquine and hydroxychloroquine are related molecules), which blocks one of the final steps in autophagy. Other drug candidates, directed at earlier components of the autophagy machinery, are in the pipeline. Early trials of hydroxychloroquine demonstrated only modest results, but there has been a resurgence of interest in combinatorial treatment approaches after evidence from animal models that, if signalling mediated downstream of mutant Ras by the enzyme MAP kinase is inhibited, pancreatic cancer cells become strongly dependent on autophagy for their survival<sup>14,15</sup>.

Yamamoto and colleagues’ work will almost certainly lead to further additions to the compendium of autophagy-targeted clinical trials of pancreatic cancer treatments. Discoveries in the fields of autophagy and immunotherapy were, respectively, recognized by the Nobel Prize in Physiology or Medicine in 2016 and 2018. This new finding represents

an unprecedented opportunity for the convergence of these two areas of study, in efforts to improve therapies for pancreatic cancer.

**Anirban Maitra** is in the Sheikh Ahmed Center for Pancreatic Cancer Research, University of Texas MD Anderson Cancer Center, Houston, Texas 77005, USA.

e-mail: amaitra@mdanderson.org

1. Yamamoto, K. *et al.* *Nature* **581**, 100–105 (2020).
2. Wolchok, J. *Cell* **175**, 1452–1454 (2018).
3. O’Reilly, E. M. *et al.* *JAMA Oncol.* **5**, 1431–1438 (2019).
4. Balachandran, V. P., Beatty, G. L. & Dougan, S. K.

- Gastroenterology* **156**, 2056–2072 (2019).
5. Yao, W., Maitra, A. & Ying, H. *EBioMedicine* **53**, 102655 (2020).
6. Whittle, M. C. & Hingorani, S. R. *Gastroenterology* **156**, 2085–2096 (2019).
7. Pommier, A. *et al.* *Science* **360**, ea404908 (2018).
8. Gettinger, S. *et al.* *Cancer Discov.* **7**, 1420–1435 (2017).
9. Grasso, C. S. *et al.* *Cancer Discov.* **8**, 730–749 (2018).
10. Waddell, N. *et al.* *Nature* **518**, 495–501 (2015).
11. Ryschich, E. *et al.* *Clin. Cancer Res.* **11**, 498–504 (2005).
12. Santana-Codina, N., Mancias, J. D. & Kimmelman, A. C. *Annu. Rev. Cancer Biol.* **1**, 19–39 (2017).
13. Yang, S. *et al.* *Genes Dev.* **25**, 717–729 (2011).
14. Bryant, K. L. *et al.* *Nature Med.* **25**, 628–640 (2019).
15. Kinsey, C. G. *et al.* *Nature Med.* **25**, 620–627 (2019).

This article was published online on 22 April 2020.

## Precision measurements

# Mass spectrometry for future atomic clocks

**Marianna S. Safronova**

Highly charged ions could form the basis of the next generation of ultra-precise clocks, using electronic transitions in the ions as the ‘pendulum’. An ingenious method for characterizing such transitions has been reported. **See p.42**

Atomic clocks, which use transitions between the energy levels of electrons in atoms as a reference for their timekeeping mechanism, are the world’s most accurate clocks — they will not lose one second during the lifetime of the Universe<sup>1</sup>. This means that they can be used in ultra-precise measurements to probe some of the fundamental postulates of modern physics. Clocks based on highly charged ions (HCIs; atoms from which many electrons have been removed) are predicted to have even more sensitivity in these investigations<sup>2</sup>. However, the development of such clocks

**“The authors used Einstein’s principle of energy–mass equivalence to convert a mass measurement into an energy measurement.”**

is hampered by the difficulty of detecting suitable transitions in HCIs.

On page 42, Schüssler *et al.*<sup>3</sup> report that they have measured a long-lived, excited electronic state in a highly charged rhenium ion using the mass difference of the ion in its ground and excited states. This non-destructive, direct determination of an electronic excitation in an HCI will aid the discovery of HCI transitions that would be suitable for use in a clock.

To build a clock, one needs a periodic event whose frequency acts as a reference for

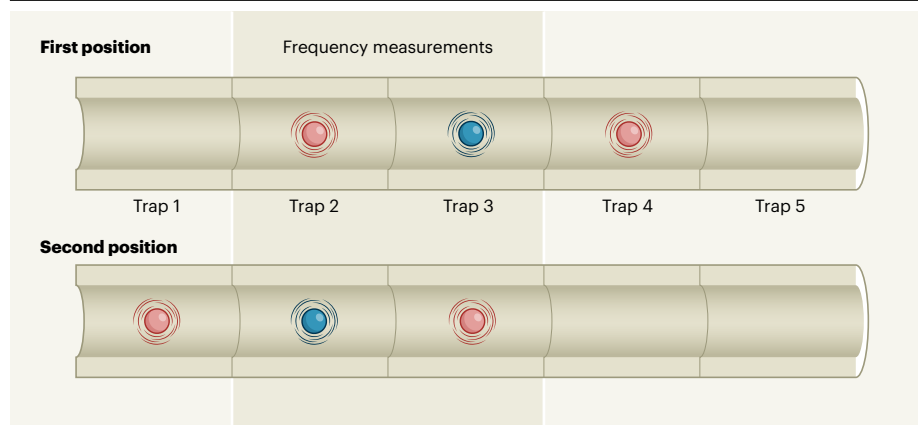
timekeeping. Electronic transitions in atoms are perfect natural oscillators for this purpose. An ultra-stable laser must be tuned to the exact frequency of the atomic transition to drive the oscillation, much as a musical instrument must be tuned to produce the right tone.

Can just any atomic transition be used? No — suitable transitions are hard to come by. The best transitions start from the lowest energy state of an atom (the ground state) and must end up in a long-lived (metastable) excited state. The energy needed to stimulate the transition must also be within the range of tabletop-laser technologies.

Moreover, the atoms must be held in traps, so that their motion is almost completely frozen — in other words, the operation of atomic clocks requires precision manipulation of quantum systems. For this reason, currently available clocks use transitions either in electrically neutral atoms or in ions produced by removing one electron from an atom, because these systems are the most amenable to precision quantum control.

Substantial advances have been made in studies of HCIs, and all the technologies required to make a clock using such an ion were demonstrated only this year<sup>4</sup>. However, progress is hindered by the difficulty in using conventional atomic spectroscopy to identify and measure transitions suitable for use in clocks — the characteristics of such transitions mean that they are, by definition, very weak (the probability of the transition occurring





**Figure 1 | Mass measurements of a highly charged ion.** Schüssler *et al.*<sup>3</sup> used an instrument called PENTATRAP, which consists of five stacked ion traps, to determine how the mass of the rhenium ion  $\text{Re}^{29+}$  differs in its ground state (blue ions) and in a metastable excited state (red ions). The authors captured three ions in their first position (traps 2–4) at 4 kelvin, and simultaneously measured the cyclotron frequency – the frequency of an ion’s motion in a magnetic field – of the ions in traps 2 and 3. All three ions were then shifted to their second position, so that the ions in traps 2 and 3 had different states from the ones that were trapped in the first position. The cyclotron frequencies of the ions in those traps were measured, and the whole sequence was repeated many times. From the ratio of cyclotron frequencies of ions in the two different states, the authors determined the associated difference in ion mass, and thus the change of energy that occurs when a  $\text{Re}^{29+}$  ion transitions between states. Such a measurement would be difficult using conventional atomic spectroscopy.

is small). Schüssler *et al.* therefore used a completely different and ingenious method to measure the energy change that occurs during a weak transition in a highly charged rhenium ion ( $\text{Re}^{29+}$ ): they used Einstein’s famous principle of energy–mass equivalence ( $E = mc^2$ ) to convert a mass measurement into an energy measurement.

The basic idea is to trap a single ion in a Penning trap, a device that confines charged particles using magnetic and electric fields. The mass of an ion in a Penning trap can be determined by measuring the frequency of the ion’s motion in a magnetic field (the cyclotron frequency). The binding energy of an atom or ion – the energy required to break the atom into its free electrons and a nucleus – is different in an excited metastable state from that in the ground state. The mass therefore also changes, which, in turn, alters the cyclotron frequency.

In their experiments, Schüssler *et al.* measured the ratio ( $R$ ) of the cyclotron frequency of  $\text{Re}^{29+}$  in the ground state and the metastable state. Because the difference in energy of the two states in  $\text{Re}^{29+}$  is extremely small compared with the total energy of the ion, the precision of the measurement needs to be extraordinarily high. The authors measured  $R$  to a precision of  $10^{-11}$ , using a device known as PENTATRAP.

PENTATRAP consists of a stack of five Penning traps cooled to a temperature of 4 kelvin (Fig. 1). Traps 2 and 3 are used to measure cyclotron frequencies, whereas traps 1 and 4 are used to store ions. Trap 5 was not used in the current experiments, but will allow monitoring of fluctuations in the magnetic field and

in other experimental variables in the future.

The authors loaded three ions into the innermost traps, so that the ions in traps 2 and 4 were in the same state (either the metastable state or the ground state), and the ion in trap 3 was in the alternative state. First, they measured the cyclotron frequencies of the ions in traps 2 and 3 simultaneously. They then moved the three ions up by one trap, effectively swapping the states of the ions in traps 2 and 3 (the states of the ions did not change, only their positions; Fig. 1), and simultaneously measured the cyclotron frequencies of those ions. The three ions were moved back down by one trap, and the sequence began again. Overall, the electronic states in traps 2 and 3 were repeatedly swapped, and simultaneous measurements were taken after each swap.

This experimental procedure, combined with the design of the PENTATRAP device, suppresses the effect of magnetic-field variations on  $R$ , thus allowing  $R$  to be determined with high accuracy. The energy difference between the ground and excited states can then be calculated using  $R$  and the ion mass in a variant of Einstein’s equation; the actual mass of the ion needs to be known only to a precision of  $10^{-4}$ .

This first demonstration of the method opens up exciting possibilities for measuring the transition energies in HCLs that are difficult to measure using conventional approaches. Moreover, the energy change measured by Schüssler and colleagues is in excellent agreement with that predicted from the authors’ advanced theoretical calculations. This agreement demonstrates that theory can be used to predict the transition energies in HCLs, thereby

facilitating the discovery of more transitions.

The transition energy measured in the current work corresponds to a frequency that lies outside the range of lasers that can be used in a clock. However, the authors note that it should be possible to use their method to measure transitions that have lower frequencies suitable for clock development in the near future.

Clocks based on HCL transitions are particularly attractive because they could be used in stringent tests that are sensitive enough to detect physics beyond the standard model of particles and interactions – such as variations of fundamental physical constants and violations of Lorentz invariance<sup>2</sup> (a cornerstone of physics that acts as the mathematical foundation for Einstein’s special theory of relativity). Such clocks would also be particularly sensitive to the effects of ultralight dark matter<sup>2,5</sup>, one of the candidates for the ‘missing’ matter in the Universe. Tremendous progress in the control of HCLs has been made in the past few years<sup>2,4,6</sup>, paving the way towards these applications. The precision mass spectrometry enabled by PENTATRAP will also have other valuable applications<sup>7</sup>, such as in tests of the energy–mass equivalence principle, experimental determinations of the upper limits of the mass of neutrino particles, and tests of quantum electrodynamics, the theory that describes the interactions between particles and light.

**Marianna S. Safronova** is in the Department of Physics and Astronomy, University of Delaware, Newark, Delaware 19716, USA.  
e-mail: msafrono@udel.edu

1. Brewer, S. M. *et al.* *Phys. Rev. Lett.* **123**, 033201 (2019).
2. Kozlov, M. G., Safronova, M. S., Crespo López-Urrutia, J. R. & Schmidt, P. O. *Rev. Mod. Phys.* **90**, 045005 (2018).
3. Schüssler, R. X. *et al.* *Nature* **581**, 42–46 (2020).
4. Micke, P. *et al.* *Nature* **578**, 60–65 (2020).
5. Safronova, M. S. *et al.* *Rev. Mod. Phys.* **90**, 025008 (2018).
6. Schmöger, L. *et al.* *Science* **347**, 1233–1236 (2015).
7. Dilling, J., Blaum, K., Brodeur, M. & Eliseev, S. *Annu. Rev. Nucl. Part. Sci.* **68**, 45–74 (2018).

# Laser spectroscopy of pionic helium atoms

<https://doi.org/10.1038/s41586-020-2240-x>

Received: 25 September 2019

Accepted: 10 March 2020

Published online: 6 May 2020

Masaki Hori<sup>1</sup>✉, Hossein Aghai-Khozani<sup>1,4</sup>, Anna Sótér<sup>1,5</sup>, Andreas Dax<sup>2</sup> & Daniel Barna<sup>3,6</sup>

Charged pions<sup>1</sup> are the lightest and longest-lived mesons. Mesonic atoms are formed when an orbital electron in an atom is replaced by a negatively charged meson. Laser spectroscopy of these atoms should permit the mass and other properties of the meson to be determined with high precision and could place upper limits on exotic forces involving mesons (as has been done in other experiments on antiprotons<sup>2–9</sup>). Determining the mass of the  $\pi^-$  meson in particular could help to place direct experimental constraints on the mass of the muon antineutrino<sup>10–13</sup>. However, laser excitations of mesonic atoms have not been previously achieved because of the small number of atoms that can be synthesized and their typically short (less than one picosecond) lifetimes against absorption of the mesons into the nuclei<sup>1</sup>. Metastable pionic helium ( $\pi^4\text{He}^+$ ) is a hypothetical<sup>14–16</sup> three-body atom composed of a helium-4 nucleus, an electron and a  $\pi^-$  occupying a Rydberg state of large principal ( $n \approx 16$ ) and orbital angular momentum ( $l \approx n - 1$ ) quantum numbers. The  $\pi^4\text{He}^+$  atom is predicted to have an anomalously long nanosecond-scale lifetime, which could allow laser spectroscopy to be carried out<sup>17</sup>. Its atomic structure is unique owing to the absence of hyperfine interactions<sup>18,19</sup> between the spin-0  $\pi^-$  and the  $^4\text{He}$  nucleus. Here we synthesize  $\pi^4\text{He}^+$  in a superfluid-helium target and excite the transition  $(n, l) = (17, 16) \rightarrow (17, 15)$  of the  $\pi^-$ -occupied  $\pi^4\text{He}^+$  orbital at a near-infrared resonance frequency of 183,760 gigahertz. The laser initiates electromagnetic cascade processes that end with the nucleus absorbing the  $\pi^-$  and undergoing fission<sup>20,21</sup>. The detection of emerging neutron, proton and deuteron fragments signals the laser-induced resonance in the atom, thereby confirming the presence of  $\pi^4\text{He}^+$ . This work enables the use of the experimental techniques of quantum optics to study a meson.

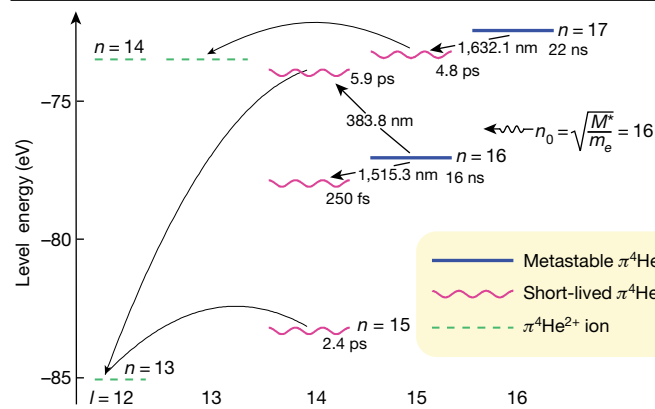
A meson consists of a valence quark and an antiquark that are bound by the strong interaction. Following the 1932 discovery of the neutron, the existence of mesons was predicted<sup>22</sup> in 1935 as the quanta associated with the hypothetical field that gives rise to the strong force binding protons and neutrons into atomic nuclei. From the apparent femtometer-scale range of the nuclear force, which was assumed to be mediated by the exchange<sup>23</sup> of mesons, the new particle was predicted to be 200 times more massive than the electron. In 1947, charged pions having roughly the expected mass became the first mesons to be discovered<sup>24</sup> among the secondary particles produced by cosmic rays. In the same year, Fermi and Teller suggested<sup>25</sup> that various exotic atoms and ions may be formed when one or more electrons in a normal atom are replaced by a heavy, negatively charged particle such as a  $\pi^-$ . Whereas isolated  $\pi^-$  mesons decay via the weak interaction with a lifetime<sup>13</sup> of  $\tau = 26$  ns into a negatively charged muon ( $\mu^-$ ) and a muon antineutrino ( $\bar{\nu}_\mu$ ),  $\pi^-$  mesons that occupy excited atomic states typically undergo rapid ( $\leq 1$  ps) electromagnetic cascade processes<sup>25,26</sup> and deexcite to  $s$ ,  $p$  or  $d$  atomic states. As these low- $l$  orbitals have a large overlap with the atomic nucleus, the  $\pi^-$  is absorbed by the nucleus via the strong interaction. Metastable  $\pi^4\text{He}^+$  atoms have nevertheless been predicted with a lifetime more than 1,000 times longer than those of other pionic atoms. The long

lifetime arises because the large binding energy of helium ( $I_0 = 25$  eV) and the high multipolarity of the transition ( $\Delta l \geq 3$ ) inhibit the deexcitation of the  $\pi^4\text{He}^+$  by Auger emission of the  $1s$  electron<sup>14–16</sup>. The remaining electron protects the stability of the  $\pi^-$  orbital and prevents Stark mixing with the low- $l$  orbitals during the many collisions with normal helium atoms that the  $\pi^4\text{He}^+$  makes during its lifetime. This extreme longevity (for a mesonic atom) makes  $\pi^4\text{He}^+$  a favourable candidate for laser excitation<sup>17</sup>.

For context, laser spectroscopy has previously been achieved for three-body antiprotonic helium ( $\bar{p}\text{He}^+ \equiv \bar{p} + \text{He}^{2+} + e^-$ ) in which an antiproton  $\bar{p}$  occupies an  $n \approx 38$  orbital with a microsecond-scale lifetime<sup>5–7</sup>. This allowed the determination<sup>6</sup> of the antiproton-to-electron mass ratio with a fractional precision of  $8 \times 10^{-10}$ . On the other hand, the smaller mass of  $\pi^-$  (seven times smaller than that of  $\bar{p}$ ) means that  $\pi^4\text{He}^+$  is much less stable and amenable to direct detection than  $\bar{p}\text{He}^+$ , which resembles a diatomic molecule<sup>15,16</sup>. Consequently, no atomic lines of  $\pi^4\text{He}^+$  have been observed so far. The existence of this atom has been indirectly inferred from experiments<sup>27–31</sup> in which the lifetime against nuclear absorption of  $\pi^-$  mesons brought to rest in helium targets was estimated to be  $\tau \approx 300$  ps–7 ns. These values cannot be easily interpreted because they differ from the expected Auger lifetimes of the  $\pi^4\text{He}^+$  states<sup>15–17</sup>.

<sup>1</sup>Max-Planck-Institut für Quantenoptik, Garching, Germany. <sup>2</sup>Paul Scherrer Institut, Villigen, Switzerland. <sup>3</sup>CERN, Geneva, Switzerland. <sup>4</sup>Present address: McKinsey and Company, Munich, Germany. <sup>5</sup>Present address: ETH Zürich, IPA, Zurich, Switzerland. <sup>6</sup>Present address: Institute for Particle and Nuclear Physics, Wigner Research Centre for Physics, Budapest, Hungary.

✉e-mail: Masaki.Hori@mpq.mpg.de



**Fig. 1 | Energy level diagram of  $\pi^4\text{He}^+$ .** On the left-hand scale the theoretical absolute energy of each  $\pi^4\text{He}^+$  state with principal and angular momentum quantum numbers ( $n, l$ ) is plotted relative to the three-body breakup threshold. The wavy lines indicate Auger-dominated states with picosecond-scale lifetimes, and the solid lines show metastable levels with theoretical lifetimes greater than 10 ns. The indicated lifetimes were calculated<sup>17</sup> as the reciprocal of the sum of the Auger and radiative decay rates of each state and the intrinsic  $3.8 \times 10^7 \text{ s}^{-1}$  weak decay rate of  $\pi^-$ . Dashed lines indicate the two-body  $\pi^4\text{He}^{2+}$  ionic states composed of a helium nucleus and a  $\pi^-$ , and curved arrows show some Auger transitions with multipolarity  $\Delta l = 2$ . We searched for the laser transitions indicated by the straight arrows.

The spectroscopic signal that would provide unambiguous evidence for the existence of  $\pi^4\text{He}^+$  can be observed only if a substantial population of  $\pi^-$  mesons were to occupy the states interrogated by the laser<sup>17</sup>. Nothing is experimentally known about the quantum numbers of these populated states, but we assumed<sup>14–16</sup> that  $n$  would be distributed around  $n_0 \equiv \sqrt{M^*/m_e} \approx 16$ , where  $M^*$  denotes the reduced mass of the pion–nucleus system and  $m_e$  is the electron mass. This corresponds to a highly excited  $\pi^-$  orbital with the same radius and binding energy as those of the replaced electron. Figure 1 shows the energy level diagram of  $\pi^4\text{He}^+$ , obtained from three-body quantum electrodynamics calculations<sup>17</sup> that include relativistic and radiative recoil corrections of order  $m_e\alpha^4$ , where  $\alpha$  denotes the fine-structure constant. The complex-coordinate-rotation method allows the energies of states with even large Auger widths ( $\Gamma_A \approx 30 \text{ GHz}$ ) to be calculated<sup>5,17</sup> with a relative precision of  $< 10^{-8}$ . The precision of the theoretical transition frequencies  $\nu_{th}$  are limited by the  $10^{-6}$ -scale experimental uncertainty<sup>13,17</sup> of the  $\pi^-$  mass. The effects of collisions<sup>32,33</sup> with helium atoms, which may destroy some  $\pi^4\text{He}^+$  states, are difficult to estimate and were not included in the calculated lifetimes shown in Fig. 1.

We used sub-nanosecond laser pulses to induce a transition from a  $\pi^4\text{He}^+$  state with a nanosecond-scale lifetime to a state with a picosecond-scale lifetime against Auger decay<sup>17</sup>. The two-body pionic helium ion ( $\pi^4\text{He}^{2+} \equiv \pi^- + {}^4\text{He}^{2+}$ ) that remained after Auger emission was destroyed in atomic collisions that induced Stark mixing between Rydberg and low- $l$  orbitals. The resonance condition between the  $\pi^4\text{He}^+$  and the laser was detected as a peak in the total production rate of neutrons, protons and deuterons that emerged from the resulting nuclear absorption. We resolved this signal, which was superimposed on a background of  $\pi^4\text{He}^+$  atoms that spontaneously decayed with a lifetime<sup>17,31</sup> of  $\sim 7 \text{ ns}$ , by measuring the nuclear fission products using scintillation counters with nanosecond-scale timing resolutions. Larger dispersions in the arrival times of the  $\pi^-$  and laser pulses at the experimental target, or fission products at the detectors, would cause the signal to broaden and become indistinguishable from the background, which was up to  $10^4$  times larger.

The experiment was carried out at the  $\pi\text{E5}$  beamline (see Fig. 2a) of the 590-MeV ring cyclotron facility<sup>34,35</sup> at the Paul Scherrer Institute (PSI), which provided a  $\pi^-$  beam with momenta between  $p = 83 \text{ MeV } c^{-1}$

and  $87 \text{ MeV } c^{-1}$  with the highest average intensity  $N_\pi = 2\text{--}3 \times 10^7 \text{ s}^{-1}$  available worldwide. The large contamination ( $> 3 \times 10^9 \text{ s}^{-1}$ ) of electrons in the beam was removed using a 550-kV  $E \times B$  Wien filter. The purified  $\pi^-$  beam arrived at the  ${}^4\text{He}$  target at timings spaced by  $\Delta t = 19.75 \text{ ns}$ , which arose from the  $f_a = 50.63 \text{ MHz}$  accelerating radiofrequency of the cyclotron. Each radiofrequency cycle contained on average  $N_\pi/f_a = 0.4\text{--}0.6$   $\pi^-$  mesons. The  $\pi^-$  mesons were distinguished (see Fig. 3a) from the residual  $e^-$  and  $\mu^-$  by their different time of flight from the graphite target<sup>36</sup> where the particles were created, as well as by the characteristic energy deposition  $\Delta E$  of the particles traversing a segmented scintillation counter positioned upstream of the  ${}^4\text{He}$  target.

Around 2.3% of the  $\pi^-$  mesons that came to rest in the superfluid- ${}^4\text{He}$  target (length 150 mm, diameter 42 mm, temperature  $T = 1.7 \text{ K}$ ; Fig. 2b) were assumed<sup>31</sup> to form  $\pi^4\text{He}^+$ . This implies a production rate of  $> 3 \times 10^5 \text{ s}^{-1}$ , which ensures a  $10^{-3}$ -scale probability of coincidence<sup>17</sup> between the laser pulse and the  $\pi^4\text{He}^+$  atom. A laser beam of diameter  $d = 25 \text{ mm}$ , pulse energy  $E_p = 10 \text{ mJ}$ , pulse length  $\Delta t = 800 \text{ ps}$ , repetition rate  $f_r = 80.1 \text{ Hz}$  and wavelengths  $\lambda = 1,515\text{--}1,634 \text{ nm}$  entered the target chamber through a fused-silica window. The absence of bubbles in the superfluid  ${}^4\text{He}$  allowed the laser light to reach the  $\pi^4\text{He}^+$  without considerable scattering.

The 1–3 neutrons and the proton or deuteron that typically emerged from a  $\pi^-$  absorption followed anticollinear<sup>20,21</sup> trajectories and had kinetic energies distributed between a few megaelectronvolts and  $E = 60\text{--}80 \text{ MeV}$ . The arrival times  $t_a$  and energy depositions  $\Delta E$  of these fission products were measured (Fig. 3b) by an array of 140 plastic scintillation counters covering a solid angle of  $2\pi$  steradians as seen from the target. This high segmentation and the high-speed waveform digitization<sup>37</sup> of the signals allowed us to resolve individual  $\pi\text{He}^+$  events despite the high intensity of the  $\pi^-$  beam. Each  $40 \times 35 \times 34 \text{ mm}^3$  counter in the array provided a  $< 10\%$  detection efficiency to signal neutrons<sup>17,38</sup> of kinetic energy  $E \geq 25 \text{ MeV}$ . The background electrons that arose from the beam or the decay of  $\mu^-$  in the experimental apparatus deposited an average energy of  $\Delta E = 6\text{--}8 \text{ MeV}$  in a counter. Most of these were removed by rejecting events with  $\Delta E < 20\text{--}25 \text{ MeV}$ .

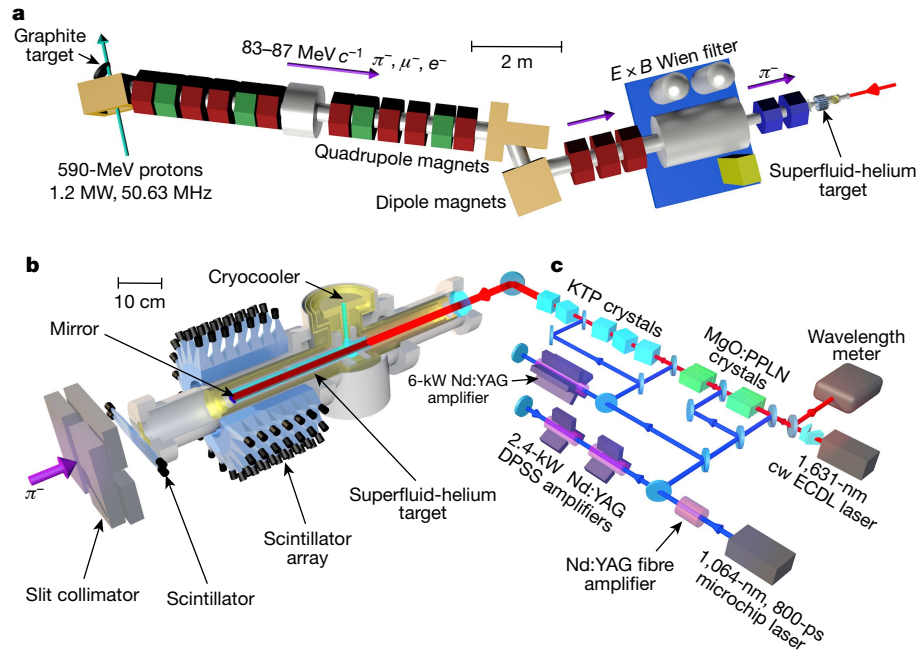
The histogram in Fig. 3c shows the distribution of the arrival times of events at the 140 scintillators measured in this way without laser irradiation. The  $\pi^-$  arrivals at  $t = 0$  and  $t = 19.75 \text{ ns}$  produced peaks representing the more than 97% of  $\pi^-$  mesons that underwent immediate nuclear absorption. The remaining  $(2.1 \pm 0.7)\%$  fraction gave a continuous spectrum that decayed with a lifetime of  $\tau = 7 \pm 2 \text{ ns}$  in the intervals between  $\pi^-$  arrivals. This roughly agrees with the results of a Monte Carlo simulation<sup>17</sup> for the signal expected from  $\pi\text{He}^+$  and with a previous experiment<sup>31</sup> that used a liquid helium target.

The laser pulses irradiated the target at  $t = 9 \text{ ns}$  after the arrival of the  $\pi^-$  mesons with a typical timing jitter of  $\Delta t \leq 1 \text{ ns}$ . For this, we developed an injection-seeded, optical parametric generation (OPG) and amplification (OPA) laser system<sup>39</sup> and a pump laser based on an amplified microchip laser (Fig. 2c), which was precisely synchronized with the radiofrequency of the cyclotron.

We initially assumed that the states  $n = 16$  were the most populated and searched for the transition  $(n, l) = (16, 15) \rightarrow (17, 14)$ . A continuous-wave pulse-amplified dye and Ti:sapphire hybrid laser was used to irradiate the  ${}^4\text{He}$  target with 700-ps-long laser pulses of energy  $E_p = 7\text{--}8 \text{ mJ}$  and diameter  $d = 20 \text{ mm}$ . The laser was scanned over a 200-GHz region around the calculated<sup>17</sup> transition frequency of  $\nu_{th} = 781052.6(2.0)_{th} \text{ GHz}$ . The theoretical ( $th$ ) uncertainty of 2.0 GHz mostly arises from the uncertainty of the  $\pi^-$  mass<sup>13</sup> used in the calculation. No significant signal was observed. Subsequent calculations revealed that the coupling of the daughter state  $(17, 14)$  to an electronically excited  $\pi^4\text{He}^+$  state gives rise to large scalar and tensor polarizabilities of  $4 \times 10^4$  and 70 atomic units, respectively<sup>32</sup>. This can destabilize state  $(17, 14)$  against atomic collisions and prevent the detection of the resonance.

We next searched for  $(16, 15) \rightarrow (16, 14)$  at  $\lambda = 1,515.3 \text{ nm}$  using the OPG-OPA laser. The 250-fs lifetime<sup>17</sup> of the  $(16, 14)$  state implies a large





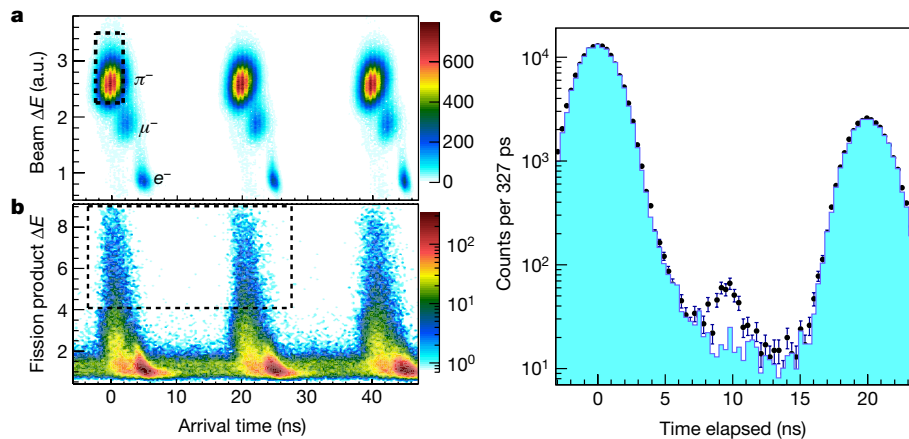
**Fig. 2 | Experimental layout.** **a**, Schematic layout of the  $\pi^-$  beamline<sup>34,35</sup>. The  $\pi^-$  mesons are produced by allowing a proton beam of power  $P=1.2$  MW and repetition rate  $f_a=50.63$  MHz to traverse a graphite target<sup>36</sup>.  $\pi^-$  mesons with momentum  $p=83\text{--}87\text{ MeV } c^{-1}$  and intensity  $N_\pi=2\text{--}3 \times 10^7\text{ s}^{-1}$  are transported by a magnetic beamline to the position of a superfluid-helium target. An  $E \times B$  Wien filter removes the  $e^-$  contamination from the beam. **b**, Layout of the experimental target. The  $\pi^-$  beam traverses a segmented scintillation counter before coming to rest in the helium target, thereby producing  $\pi^4\text{He}^+$  atoms. The atoms are irradiated with laser pulses with length  $\Delta t=800$  ps, energy  $E_p=10$  mJ, diameter  $d=25$  mm and wavelength  $\lambda=1,515\text{--}1,634$  nm. The proton, neutron

and deuteron fragments<sup>20,21</sup> that emerge from the  $\pi^-$  absorption in the helium nuclei are detected by an array of 140 plastic scintillation counters surrounding the target. **c**, Schematic layout of the laser system (not to scale), which is based on a continuous-wave (cw) external-cavity diode laser (ECDL). OPG<sup>39</sup> and OPA are carried out in magnesium-doped periodically poled lithium niobate (MgO:PPLN) and potassium titanyl phosphate (KTP) crystals to produce the laser pulses that irradiate the  $\pi^4\text{He}^+$  atoms. The laser beams used to pump the crystals are produced by amplifying the output of a microchip laser of  $\lambda=1,064$  nm in neodymium-doped yttrium aluminium garnet (Nd:YAG) fibre and rod crystals (see text).

width of  $\Gamma_A=640$  GHz for the resonance. Data accumulated over 300 h corresponding to more than  $6 \times 10^7$  detected  $\pi^-$  arrivals showed no substantial signal. The reason for this is not understood, but atomic collisions may destroy the  $\pi^-$  population occupying state (16, 15), as

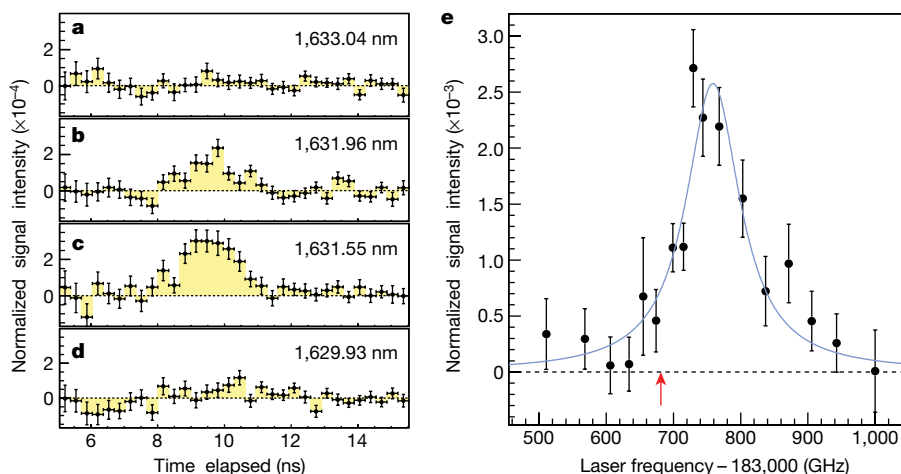
observed<sup>40</sup> for several  $\bar{p}\text{He}^+$  states. Alternatively, it is possible that a very small  $\pi^-$  population is captured into (16, 15).

We shifted our search to the transition (17, 16)  $\rightarrow$  (17, 15) in the  $n=17$  region. This resonance with an amplitude<sup>17</sup> of  $M=1.4$  atomic units



**Fig. 3 | Detection of  $\pi^-$  arrivals and nuclear absorptions in the helium target, and time spectra of the emerging fission fragments measured with and without laser irradiation.** **a**, Contour plot showing the correlation between the arrival times  $t_a$  and energy depositions  $\Delta E$  of particles measured by a scintillation counter positioned at the entrance of the experimental target. The particles arrived at timings spaced by  $\Delta t=19.75$  ns. The particle type ( $\pi^-$ ,  $\mu^-$  or  $e^-$ ) was identified, and the  $\pi^-$  events in the dashed rectangle were selected. a.u., arbitrary units. **b**,  $t_a\text{--}\Delta E$  contour plot of emerging fission fragments that strike the scintillator array following the  $\pi^-$  absorption by helium nuclei.

Background electrons of  $\Delta E < 20\text{--}25$  MeV and fission products of low velocity were removed by accepting events with energy depositions between  $\Delta E=20\text{--}25$  MeV at the lower limit and  $\Delta E \geq 60$  MeV at the higher limit in the dashed rectangle. **c**, Time spectra of fission fragments measured with (filled circles with error bars) and without (histogram) laser pulses of wavelength  $\lambda \approx 1,631.4$  nm irradiating the target at  $t=9$  ns. The peak at  $t=9$  ns in the former spectrum corresponds to the resonance signal of the transition  $(n, l) = (17, 16) \rightarrow (17, 15)$  of metastable  $\pi^4\text{He}^+$ . Error bars, 1 s.d.



**Fig. 4 | Laser resonance signals of the transition  $(n, l) = (17, 16) \rightarrow (17, 15)$  in metastable  $\pi^4\text{He}^+$  atoms.** **a–d**, Normalized time spectra of the resonance signal measured at four laser wavelengths. The plots are obtained by taking the difference between the timing distributions of  $\pi^-$  absorption in  $^4\text{He}$  nuclei measured with and without laser irradiation. When the laser is tuned to a wavelength of  $\lambda \approx 1,631.4$  nm, the transition  $(17, 16) \rightarrow (17, 15)$  is excited and a peak is generated at  $t = 9.5$  ns at the rate of  $\pi^-$  absorptions. Error bars, 1 s.d. **e**, Profile of the resonance measured by scanning the optical frequency of the laser over a

500-GHz-wide region and plotting the normalized number of counts under the peaks. Black filled circles indicate experimental data points with 1-s.d. error bars, the blue line is the best fit of a double Lorentzian function, and the red arrow indicates the position of the spin-averaged transition frequency obtained by a three-body quantum electrodynamics calculation<sup>17</sup>. The systematic uncertainty<sup>39,42</sup> of the laser frequency arising from the OPG and OPA processes is  $\pm 3$  GHz.

should readily saturate at our laser fluence of  $\varepsilon = 2 \text{ mJ cm}^{-2}$ . The time spectrum indicated by the filled circles in Fig. 3c represents  $2.5 \times 10^7$   $\pi^-$  arrivals detected over 100 h, with the laser wavelength tuned to  $\lambda \approx 1,631.4$  nm. A sharp peak was observed at  $t = 9$  ns containing 300 events with a signal-to-noise ratio of 4 and a statistical significance of  $>7$  s.d., in coincidence with the laser pulses. The width  $\Delta t = 2$  ns of the peak is commensurate with the dispersions in the times of flight of the fission fragments arriving at the scintillator array. The measured detection rate of  $3 \text{ h}^{-1}$  of resonant  $\pi^4\text{He}^+$  is compatible with the production rate of  $>3 \times 10^5 \text{ s}^{-1}$  of the atoms and with Monte Carlo simulations<sup>17</sup> that assume that most of the metastable population lies in the  $(17, 16)$  state. As expected, detuning the laser frequency away from the  $\pi^4\text{He}^+$  resonance (Fig. 4a–d) caused the signal to gradually decrease and finally disappear.

By scanning the laser frequency over a 500-GHz-wide region and plotting the relative number of counts under the laser-induced peak, we obtained the resonance profile of Fig. 4e. The statistical uncertainty of the signal is indicated by the error bars. The  $\sim 100$ -GHz width of the resonance roughly agrees with a convolution of the calculated<sup>17</sup> Auger width  $\Gamma_A = 33$  GHz of state  $(17, 15)$ , collisional<sup>33</sup> and power broadening ( $\sim 50$  GHz) and the linewidth ( $\sim 10$  GHz) of the seeded narrowband component of the laser beam. Additional width may be due to atomic collisions that shorten<sup>32,40</sup> the lifetime of  $(17, 15)$ .

The 3.0-GHz spacing<sup>17</sup> between the fine-structure sublines that arise from the interaction between the electron spin and the orbital angular momentum of  $\pi^-$  is small compared to the 33-GHz natural width of the resonance. The best fit (blue curve) of two overlapping Lorentzian functions that take the sublines into account have a reduced  $\chi^2$  value of 1.0. Lorentzian lineshapes are expected<sup>17</sup> for resonances that are homogeneously broadened by the natural width, binary atomic collisions and power-broadening effects. The resonance centroid was determined as  $\nu_{\text{exp}} = 183760(6)_{\text{stat}}(6)_{\text{syst}}$  GHz. The statistical ('stat') uncertainty of 6 GHz is due to the finite number of detected  $\pi^4\text{He}^+$  atoms. The systematic ('syst') uncertainty of 6 GHz contains contributions related to the selection of the fit function (5 GHz), the calibration of the laser frequency against the absorption lines of  $^{13}\text{C}^{16}\text{O}$  molecules<sup>41</sup> (0.8 GHz) and the uncertainty<sup>39,42</sup> related to the OPG and OPA processes (3 GHz).

The fact that  $\nu_{\text{exp}}$  is larger than the calculated value<sup>17</sup> of  $\nu_{\text{th}} = 183681.8(0.5)_{\text{th}}$  GHz by  $\Delta\nu = 78$  GHz is believed to be due to atomic

collisions that perturb the energy levels and shift the resonance frequency<sup>33</sup>, as has been observed<sup>40</sup> for  $\bar{p}\text{He}^+$ . The effect was calculated using the impact approximation of the binary collision theory of spectral lineshapes<sup>33</sup>. The gradient of the shift at a target temperature of  $T = 4$  K was predicted to be  $d\nu/d\rho = 4.4\text{--}6.5 \times 10^{-21} \text{ GHz cm}^3$ , depending on the interatomic potential and model used. At the superfluid target density,  $\rho = 2.18 \times 10^{22} \text{ cm}^{-3}$ , this corresponds to a blueshift of  $\Delta\nu = 96\text{--}142$  GHz, which roughly agrees with the above experimental result despite the complexity of the multi-body calculations.

We have measured the resonance frequency of  $\nu_{\text{exp}} = 183760(6)_{\text{stat}}(6)_{\text{syst}}$  GHz of the transition  $(17, 16) \rightarrow (17, 15)$  for  $\pi^4\text{He}^+$  produced in a superfluid-helium target, thus demonstrating laser spectroscopy of an atom containing a meson. The shift of  $\nu_{\text{exp}}$  caused by collisions must be measured before we can determine the  $\pi^-$  mass. We plan to search for resonances such as  $(17, 16) \rightarrow (16, 15)$  that are predicted to be narrower by a factor of at least  $10^3$ . By measuring  $\nu_{\text{exp}}$  at various densities of a helium gas target, the zero-density value may be determined<sup>40</sup>. Although it is difficult to predict how precisely the  $\pi^-$  mass may ultimately be determined, the natural width of some resonances implies a fractional precision of  $\leq 10^{-8}$ , if the  $\pi^4\text{He}^+$  atom could be synthesized in targets of sufficiently low density as in the  $\bar{p}\text{He}^+$  case<sup>6,7</sup>. Along with  $\text{H}_2^+$  and  $\text{HD}^+$  molecular ions<sup>5,43,44</sup> and  $\bar{p}\text{He}^+$  atoms, metastable  $\pi^4\text{He}^+$  is now among the fundamental three-body systems containing two massive particles that can be precisely studied by laser spectroscopy.

## Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-020-2240-x>.

1. Ericson, T. & Weise, W. *Pions and Nuclei* (Clarendon Press, 1988).
2. Ahmadi, M. et al. Observation of the 1S–2S transition in trapped antihydrogen. *Nature* **541**, 506–510 (2017).

3. DiSciaccia, J. et al. One-particle measurement of the antiproton magnetic moment. *Phys. Rev. Lett.* **110**, 130801 (2013).
4. Ulmer, S. et al. High-precision comparison of the antiproton-to-proton charge-to-mass ratio. *Nature* **524**, 196–199 (2015).
5. Korobov, V. I., Hilico, L. & Karr, J. -Ph.  $ma^2$ -order corrections in the hydrogen molecular ions and antiprotonic helium. *Phys. Rev. Lett.* **112**, 103003 (2014).
6. Hori, M. et al. Buffer-gas cooling of antiproton helium to 1.5 to 1.7 K, and antiproton-to-electron mass ratio. *Science* **354**, 610–614 (2016).
7. Hori, M. et al. Two-photon laser spectroscopy of antiprotonic helium and the antiproton-to-electron mass ratio. *Nature* **475**, 484–488 (2011).
8. Salumbides, E. J., Ubachs, W. & Korobov, V. I. Bounds on the fifth forces at sub-Å length scale. *J. Mol. Spectrosc.* **300**, 65–69 (2014).
9. Ficek, F. et al. Constraints on exotic spin-dependent interactions between matter and antimatter from antiprotonic helium spectroscopy. *Phys. Rev. Lett.* **120**, 183002 (2018).
10. Assamagan, K. et al. Upper limit of the muon-neutrino mass and charged-pion mass from momentum analysis of a surface muon beam. *Phys. Rev. D* **53**, 6065–6077 (1996).
11. Trassinelli, M. et al. Measurement of the charged pion mass using X-ray spectroscopy of exotic atoms. *Phys. Lett. B* **759**, 583–588 (2016).
12. Daum, M., Frosch, R. & Kettle, P.-R. The charged and neutral pion masses revisited. *Phys. Lett. B* **796**, 11–14 (2019).
13. Particle Data Group. Review of particle physics. *Chin. Phys. C* **40**, 100001 (2016).
14. Condo, G. T. On the absorption of negative pions by liquid helium. *Phys. Lett.* **9**, 65–66 (1964).
15. Russell, J. E. Metastable states of  $arr^-e^-$ ,  $aK^-e^-$ , and  $a\bar{p}e^-$  atoms. *Phys. Rev. Lett.* **23**, 63–64 (1969).
16. Russell, J. E. Structure of neutral mesonic atoms formed in liquid helium. *Phys. Rev. A* **1**, 721–734 (1970).
17. Hori, M., Sôtér, A. & Korobov, V. I. Proposed method for laser spectroscopy of pionic helium to determine the charged-pion mass. *Phys. Rev. A* **89**, 042515 (2014).
18. Koch, J. & Scheck, F. Quadrupole interaction in pionic and kaonic atoms. *Nucl. Phys. A* **340**, 221–239 (1980).
19. Trassinelli, M. & Indelicato, P. Relativistic calculations of pionic and kaonic atoms' hyperfine structure. *Phys. Rev. A* **76**, 012510 (2007).
20. Daum, E. et al. Pion absorption at rest in  $^4\text{He}$ . *Nucl. Phys. A* **589**, 553–584 (1995).
21. Cernigoi, C. et al. Energy spectra of single neutrons and charged particles emitted following the absorption of stopped negative pions in  $^4\text{He}$ . *Nucl. Phys. A* **352**, 343–354 (1981).
22. Yukawa, H. On the interaction of elementary particles. I. *Proc. Phys.-Math. Soc. Jpn.* **17**, 48–57 (1935).
23. Miller, A. I. Werner Heisenberg and the beginning of nuclear physics. *Phys. Today* **38**, 60–68 (1985).
24. Lattes, C. M. G., Muirhead, H., Occhialini, G. P. S. & Powell, C. F. Processes involving charged mesons. *Nature* **159**, 694–697 (1947).
25. Fermi, E. & Teller, E. The capture of negative mesotrons in matter. *Phys. Rev.* **72**, 399–408 (1947).
26. Gotta, D. Precision spectroscopy of light exotic atoms. *Prog. Part. Nucl. Phys.* **52**, 133–195 (2004).
27. Fetkovich, J. G. & Pewitt, E. J. Experimental study of the cascade time of negative mesons in a liquid helium bubble chamber. *Phys. Rev. Lett.* **11**, 290–293 (1963).
28. Block, M. M. et al. Moderation time for nuclear capture of negative pions in liquid  $^4\text{He}$ . *Phys. Rev. Lett.* **11**, 301–303 (1963).
29. Block, M. M., Kopelman, J. B. & Sun, C. R. Moderation and cascade time of negative pions and negative kaons in liquid helium. *Phys. Rev.* **140**, B143–B152 (1965).
30. Zaimidoroga, O. A., Sulyaev, R. M. & Tsupko-Sitnikov, V. M. Measurement of the  $\pi^-$  meson cascade transition time in gaseous  $^3\text{He}$ . *Sov. Phys. JETP* **25**, 63–64 (1967).
31. Nakamura, S. N. et al. Negative-pion trapping by a metastable state in liquid helium. *Phys. Rev. A* **45**, 6202–6208 (1992).
32. Korobov, V. I., Bekbaev, A. K., Aznabayev, D. T. & Zhaugasheva, S. A. Polarizability of the pionic helium atom. *J. Phys. B* **48**, 245006 (2015).
33. Obreshkov, B. & Bakalov, D. Collisional shift and broadening of the transition lines of pionic helium. *Phys. Rev. A* **93**, 062505 (2016).
34. Seidel, M. et al. Production of a 1.3 MW proton beam at PSI. In *Proc. IPAC10* 1309–1313 (Asian Committee for Future Accelerators, 2010).
35. Abela, R., Foroughi, F. & Renker, D. Muon beams at PSI. *Z. Phys. C* **56**, S240–S242 (1992).
36. Heidenreich, G. Carbon and beryllium targets at PSI. *Proc. AIP* **642**, 122–124 (2002).
37. Ritt, S., Dinapoli, R. & Hartmann, U. Application of the DRS chip for fast waveform digitizing. *Nucl. Instrum. Methods Phys. Res. A* **623**, 486–488 (2010).
38. Cecil, R. A., Anderson, B. D. & Madey, R. Improved predictions of neutron detection efficiency for hydrocarbon scintillators from 1 MeV to about 300 MeV. *Nucl. Instrum. Methods* **161**, 439–447 (1979).
39. Cocuzzi, M. D., Schepler, K. L. & Powers, P. E. Narrow-bandwidth, subnanosecond, infrared pulse generation in PPLN pumped by a fiber amplifier-microchip oscillator. *IEEE J. Sel. Top. Quantum Electron.* **15**, 372–376 (2009).
40. Hori, M. et al. Sub-ppm laser spectroscopy of antiprotonic helium and a CPT-violation limit on the antiprotonic charge and mass. *Phys. Rev. Lett.* **87**, 093401 (2001).
41. Swann, W. C. & Gilbert, S. L. Pressure-induced shift and broadening of 1560–1630-nm carbon monoxide wavelength-calibration lines. *J. Opt. Soc. Am. B* **19**, 2461–2467 (2002).
42. White, R. T. et al. Control of frequency chirp in nanosecond-pulsed laser spectroscopy. 3. Spectrotemporal dynamics of an injection-seeded optical parametric oscillator. *J. Opt. Soc. Am. B* **24**, 2601–2609 (2007).
43. Biesheuvel, J., Karr, J. -Ph., Hilico, L., Eikema, K. S., Ubachs, E. W. & Koelmeij, J. C. J. Probing QED and fundamental constants through laser spectroscopy of vibrational transitions in  $\text{HD}^+$ . *Nat. Commun.* **7**, 10385 (2016).
44. Alighanbari, S., Hansen, M. G., Korobov, V. I. & Schiller, S. Rotational spectroscopy of cold and trapped molecular ions in the Lamb–Dicke regime. *Nat. Physics* **14**, 555–559 (2018).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2020



## Methods

### Pion beam

A proton beam<sup>34</sup> of energy  $E = 590$  MeV and power  $P = 1.2$  MW was allowed to traverse a rotating graphite target<sup>36</sup> of thickness  $t_d = 40$  mm, thereby producing secondary  $\pi^-$ ,  $\mu^-$  and  $e^-$  (see Fig. 2a). Particles with momenta between  $p = 83$  and  $87$  MeV  $c^{-1}$  were collected by magnetic elements<sup>35</sup> and transported over a 14.2-m-long beamline to the position of the superfluid- $^4\text{He}$  target. The momentum was chosen to allow the largest possible number of  $\pi^-$  mesons to rapidly come to rest in the  $^4\text{He}$  target before they could decay while reducing the background arising from faster and slower particles that stopped in the walls of the apparatus. In our initial experiments, the high intensity ( $>3 \times 10^9$  s $^{-1}$ ) of contaminant electrons in the beam saturated the detectors and prevented the observation of  $\pi^+{}^4\text{He}^+$ . An  $E \times B$  Wien filter was therefore placed in the beam upstream of the target to divert most of the electrons towards the steel blades of a slit collimator with a thickness of 80 mm. The filter consisted of a dipole magnet and a pair of 1.67-m-long electrode plates separated by a 150-mm gap and with a voltage difference of  $\Delta V = 550$  kV. This was the highest voltage that could be applied stably without frequent breakdowns. A pair of quadrupole magnets then focused the purified  $\pi^-$  beam of intensity  $N_\pi = 2\text{--}3 \times 10^7$  s $^{-1}$  into an elliptical spot of full-width at half-maximum horizontal and vertical beam diameters of 23 mm and 15 mm, respectively. The beam diameters were determined by measuring the number of  $\pi^-$  hits on a scintillator of size  $5 \times 5 \times 3$  mm<sup>3</sup> as it was slowly moved through the beam using a motorized translation stage.

The beam traversed a plastic scintillator plate with a thickness of 4.7 mm before entering the target (see Fig. 2b). The plate was segmented into four sections of size  $20 \times 20$  mm<sup>2</sup>, each of which was read out by metal-packaged photomultiplier tubes with 8-mm-diameter bialkali photocathodes. The last three dynode stages of each photomultiplier were driven with field-effect transistors arranged in an emitter–follower configuration. This prevented the high intensity of the particle beam from strongly saturating the anode currents. The timing resolution over a 10-min interval was estimated as  $\Delta t < 0.8$  ns. Radiation damage caused the light yields of the scintillators to slowly deteriorate, and they were replaced every 10–14 days.

The contour plot in Fig. 3a shows the correlations between the arrival times  $t_a$  and energy depositions  $\Delta E$  of particles on the scintillators. The  $\pi^-$  events located in the dashed rectangle were distinguished from the residual  $\mu^-$  and  $e^-$  events by the characteristic time of flight and estimated energy deposition  $\Delta E = 2.6$  MeV for  $\pi^-$ . The  $e^-$  background was thus suppressed by  $<10^{-4}$  compared to measurements carried out without the  $E \times B$  Wien filter. Monte Carlo simulations<sup>37</sup> showed that remaining  $\mu^-$  and  $e^-$  misidentified in the analysis caused a small reduction in the signal-to-noise ratio of the laser spectroscopic signal that was negligible compared to the  $>13\%$  statistical uncertainty of the signal.

### Experimental target

The target (Fig. 2b) consisted of a 150-mm-long, 42-mm-diameter cylindrical aluminium chamber that was filled with superfluid  $^4\text{He}$  of temperature  $T = 1.7$  K. The thin ( $t_d = 500$   $\mu\text{m}$ ) lateral walls of the chamber allowed the passage of the nuclear fission products with small energy losses, which could then be detected with a high efficiency. The chamber was cooled by a 1-m-long Joule–Thompson refrigerator through which superfluid  $^4\text{He}$  was continuously pumped by a roots blower and a rotary vane pump with respective pump speeds of  $500$  m<sup>3</sup> h $^{-1}$  and  $200$  m<sup>3</sup> h $^{-1}$ . The  $\pi^-$  beam entered the chamber through a 1.6-mm-thick aluminium window and traversed a 6-mm-thick, 36-mm-diameter mirror made of fused silica before reaching the superfluid  $^4\text{He}$ . A dielectric multilayer coating of reflectivity  $>98\%$  was deposited on the downstream surface of the mirror by electron beam evaporation. Monte Carlo simulations<sup>37</sup> carried out using the GEANT4 package indicated that  $>80\%$  of the  $\pi^-$  mesons came to rest in the superfluid  $^4\text{He}$ , whereas the rest stopped in the mirror or the chamber walls.

The 25-mm-diameter laser beam of pulse energy  $E_p = 10$  mJ and wavelengths between  $\lambda = 1,515$  nm and  $1,634$  nm entered through a 35-mm-diameter fused-silica window with antireflection coating that was fixed on the opposite end of the chamber. The beam irradiated  $>60\%$  of the  $\pi^+{}^4\text{He}^+$  atoms produced in the target, retro-reflected off the mirror, and exited through the window. This helped to reduce the heating of the target by the laser beam of average power  $P = 800$  mW.

### Detectors and electronics

Each of the 140 plastic scintillation counters of size  $40 \times 35 \times 34$  mm<sup>3</sup> surrounding the target was read out by photomultipliers of the type described previously with an intrinsic timing resolution of  $\Delta t \leq 1$  ns. Variations in the times of flight of the fission products emerging from the  $\pi^-$  absorption degraded this resolution to  $\Delta t \approx 2$  ns. The segmentation reduced the pile-up probability of multiple particles striking a single counter in rapid sequence. The size of each counter was carefully chosen to provide a significant ( $<10\%$ ) detection efficiency for signal neutrons<sup>38</sup> of kinetic energy  $E \geq 25$  MeV while also allowing the rejection of most of the background electrons that arose either from the particle beam or from  $\mu^-$  decays in the experimental apparatus. An electron that traversed the full thickness of a counter corresponding to  $1/10$  radiation length deposited an average energy of  $\Delta E = 6\text{--}8$  MeV, with a distribution tail<sup>13</sup> extending to higher energies. The signals were amplified by d.c. coupled amplifiers with bandwidth  $f_b = 400$  MHz and gain  $g = 10$ . The analogue waveform of the signal from each counter was simultaneously recorded over a  $\Delta t = 83.7$  ns interval for every laser pulse arrival, using 140 waveform digitizers with sampling rates  $f_s = 3.06$  Gs s $^{-1}$ . For this, we constructed a readout system using application-specific integrated circuits (ASICs) based on switched capacitor arrays<sup>37</sup> (DRS4), sample-and-hold amplifiers and pipelined switched-capacitor analogue-to-digital converters. 40 field-programmable gate arrays continuously collected and recorded the waveform data<sup>45,46</sup> at a total acquisition rate of  $13\text{--}15$  GB h $^{-1}$  on a 24-h basis.

The waveform data were analysed using an algorithm based on constant-fraction discrimination to determine the arrival times  $t_a$  and energy depositions  $\Delta E$  in the scintillator array. A typical  $t_a$ – $\Delta E$  contour plot obtained in this way is shown in Fig. 3b. Because the detector could not identify the type of particle that traversed it, we selected the events located in the dashed rectangle, with energy depositions between  $\Delta E = 20\text{--}25$  MeV and  $\Delta E \geq 60$  MeV. This removed most of the above-mentioned background electrons with energy deposition  $\Delta E < \Delta E_r = 20\text{--}25$  MeV and low-velocity fission products that could degrade the timing resolution for detecting  $\pi^-$  absorptions. This rejection threshold  $\Delta E_r$  was selected<sup>17</sup> based on Monte Carlo simulations carried out during the design of the counters. The time spectra in Fig. 3c show the distributions of the accepted events measured with and without laser irradiation at  $t = 9$  ns (see main text). The  $\Delta t = 4$  ns widths of the peaks at  $t = 0$  and  $19.75$  ns are primarily caused by variations in the times of flight of  $\pi^-$  mesons with momenta between  $p = 83$  MeV  $c^{-1}$  and  $87$  MeV  $c^{-1}$  that arrive at the  $^4\text{He}$  target.

### Laser system

The OPG-OPA laser system was based on a continuous-wave external-cavity diode laser with an output power of  $P = 50$  mW, linewidth  $\Gamma < 1$  MHz and wavelengths between  $\lambda = 1,515$  and  $1,634$  nm (see Fig. 2c). The optical frequency was continuously measured by a Michelson interference wavelength meter and calibrated against the absorption lines<sup>41</sup> of  $^{13}\text{C}^{16}\text{O}$  molecules with a precision of  $<0.8$  GHz. This seed beam was amplified by OPG<sup>39</sup> using a pair of 40-mm-long MgO:PPLN crystals set in series, which produced 800-ps-long laser pulses of energy  $E_p = 70$   $\mu\text{J}$ . OPA to  $E_p = 10$  mJ was carried out by passing these laser pulses through five KTP crystals of lengths  $l = 20$  mm. The broadband light arising from amplified spontaneous emission (ASE) in the MgO:PPLN crystals was discarded after each amplification stage using dielectric beam splitters and irises. The ASE in the laser beam was monitored with a grating

monochromator of spectral resolution  $\Delta\lambda = 300$  pm. The linewidth<sup>39</sup> of the seeded narrowband component of the laser beam excluding the ASE was  $\sim 10$  GHz. During the OPG and OPA processes, the optical frequency  $\nu_{\text{pl}}$  of the amplified laser pulse could shift<sup>42</sup> relative to the seed laser  $\nu_{\text{cw}}$ . From relative measurements of  $\nu_{\text{pl}}$  and  $\nu_{\text{cw}}$  using a Fizeau interferometer and a grating spectrometer<sup>39</sup>, and from rough numerical calculations of the OPA process<sup>42</sup>, a conservative uncertainty of 3 GHz was ascribed to the time-averaged optical frequency of the laser pulses that reached the  $\pi^4\text{He}^+$ .

We constructed a diode-pumped solid state (DPSS) Nd:YAG laser to pump the nonlinear crystals. The laser was based on an actively Q-switched microchip laser oscillator and amplifier, which produced 800-ps-long laser pulses of  $E_p = 80$   $\mu\text{J}$  and  $\lambda = 1,064$  nm. The beam was further amplified to  $E_p = 1.5$  mJ in a Nd:YAG fibre amplifier<sup>47</sup>, and then to 50 mJ by passing the beam twice through a series of three transversely pumped DPSS Nd:YAG amplifiers with crystal rod lengths of  $l = 73$ , 73 and 126 mm, diameters  $d = 3$ , 3 and 5 mm, and peak pump powers  $P = 2.4$ , 2.4 and 6 kW, respectively. The pulsed amplifiers were mounted on a  $3 \times 1.5$  m<sup>2</sup> optical table and water-cooled by three recirculating chillers having a total refrigeration power of 4 kW. The laser was triggered at a repetition rate of  $f_r = 80.1$  Hz and a timing jitter of  $\Delta t < 1$  ns in synchronization with the  $f_a = 50.63$  MHz accelerating radiofrequency of the cyclotron. The trigger signal  $f_t$  was generated by dividing the radiofrequency  $f_a$  by a factor of 632,400 using a two-phase frequency divider. This ensured that the laser pulses and  $\pi^-$  would arrive at the experimental target with the intended relative timings. The arrival time of the laser pulses, which was monitored by an infrared-sensitive film and a photomultiplier, was found to slowly drift by 1 ns over periods of a few hours, presumably owing to thermal effects in the microchip laser. This was constantly corrected.

We developed a continuous-wave pulse-amplified dye laser and Ti:sapphire amplifier to search for the transition  $(16, 15) \rightarrow (17, 14)$  at a wavelength of  $\lambda = 383.8$  nm. A continuous-wave distributed Bragg reflector semiconductor laser produced a seed beam of power  $P = 40$  mW, linewidth  $\Gamma < 2$  MHz and  $\lambda = 767$  nm. The beam was amplified to  $P = 450$  mW using a semiconductor tapered amplifier. The laser wavelength was measured using a Fizeau wavelength meter, which was calibrated to a precision of  $< 1$  GHz using a frequency-stabilized HeNe laser. The calibration was verified by carrying out absorption spectroscopy of the D<sub>2</sub> line of a natural isotopic mixture of K. The seed beam was pulse-amplified in two dye cells, which were oriented at Brewster's angle and filled with Styryl-8 dye dissolved in ethanol. This produced 800-ps-long laser pulses of energy  $E_p = 2$  mJ. The beam was further amplified to  $E_p = 20$ –25 mJ by allowing the laser pulses to make five passes through a 20-mm-long Ti:sapphire crystal<sup>48</sup>. Frequency doubling in a beta barium borate crystal generated 700-ps-long laser pulses of  $E_p = 8$  mJ and  $\lambda = 383.8$  nm. The dye cells were pumped by a frequency-doubled, amplified microchip laser system similar to the one used in the OPG-OPA laser described above. The Ti:sapphire crystal was pumped by a separate frequency-doubled Nd:YAG laser, which generated 30-ns-long laser pulses of  $E_p = 80$ –90 mJ. This DPSS laser was based on a triangular ring cavity<sup>6</sup>.

### Resonance of $\pi^4\text{He}^+$

Time spectra of  $\pi^4\text{He}^+$  were measured at 17 settings of the laser frequency around the  $(n, l) = (17, 16) \rightarrow (17, 15)$  resonance. The relative intensity of the  $\pi^4\text{He}^+$  resonance signals (see Fig. 4a–d) was determined by taking the difference between normalized time spectra measured with and without the laser irradiation and counting the number of events under the laser-induced peak within a  $\Delta t = 3.6$ -ns-wide region centred around  $t = 9.5$  ns. This interval was chosen on the basis of a Monte Carlo simulation<sup>17</sup> that estimated the width  $\Delta t$  of the resonance signal. This width arose for several reasons: first, the fission products were emitted isotropically with kinetic energies between a few megaelectronvolts and  $E = 80$  MeV<sup>17</sup>,

whereas the flight path of the particles to the array of 140 scintillation counters varied between 40 mm and 150 mm depending on the emission angle. The resulting time-of-flight dispersion caused a broadening of  $\Delta t \approx 2$  ns depending on the energy rejection threshold  $\Delta E_r$  of the scintillators. Second, Monte Carlo simulations indicate that the  $\pi^4\text{He}^+$  atoms were distributed over an 80-mm-long volume of the experimental target<sup>17</sup>. This corresponds to a 0.5-ns transit time for a laser beam that irradiated the volume, retro-reflected off the mirror and exited the chamber. Third, the timing profile of the 1,631-nm laser pulse contained several peaks<sup>39</sup> that fluctuated from shot to shot, causing a 1-ns-scale jitter in the timing when the  $\pi^4\text{He}^+$  atoms underwent laser excitation.

The vertical error bars in the resonance profile of Fig. 4e indicate the statistical uncertainty of the signals caused by the finite numbers of resonant  $\pi^4\text{He}^+$  events. Each point represents data collected over 20–30 h. Systematic uncertainties due to changes in the choice of cuts (optimized by Monte Carlo simulations; see above) were investigated by slightly varying the selection criteria for  $\pi^-$  arrival events (Fig. 3a) and the detection of fission products by the scintillator array (Fig. 3b). The effect of this on the statistical significance of the  $\pi^4\text{He}^+$  resonance signal was small and agreed with the simulation results<sup>17</sup>, whereas the experimental value of the transition frequency  $\nu_{\text{exp}}$  remained well within the systematic uncertainties of the measurement (see main text).

### Data availability

As the datasets generated during and/or analysed during the current study are in total 20 TBytes in size consisting primarily of waveform data, they are available from the corresponding author on reasonable request.

45. Sôtér, A. et al. Segmented scintillation detectors with silicon photomultiplier readout for measuring antiproton annihilations. *Rev. Sci. Instrum.* **85**, 023302 (2014).
46. Todoroki, K. et al. Instrumentation for measurement of in-flight annihilations of 130 keV antiprotons on thin target foils. *Nucl. Instrum. Methods Phys. Res. A* **835**, 110–118 (2016).
47. Martial, I., Balembois, F., Didierjean, J. & Georges, P. Nd:YAG single-crystal fiber as high peak power amplifier of pulses below one nanosecond. *Opt. Express* **19**, 11667–11679 (2011).
48. Hori, M. & Dax, A. Chirp-corrected, nanosecond Ti:sapphire laser with 6 MHz linewidth for spectroscopy of antiprotonic helium. *Opt. Lett.* **34**, 1273–1275 (2009).

**Acknowledgements** This work was supported by the Max-Planck Gesellschaft and the European Research Council (ERC-Stg). We thank the staff members of the PSI cyclotron and beamline operations groups, the PSI laboratory for particle physics (LTP), the PSI water, cryogenic and electrical engineering groups, the CERN cryogenics laboratory and its magnet and scintillator workshops, D. Bakalov, T. W. Hänsch, R. S. Hayano, V. I. Korobov, J. Mouleyre, A. Newborough, B. Obreshkov, N. Picque, O. Pirotte, R. Pohl, A. Sakai, T. Udem and T. Yamazaki. The scintillator array and cryogenic target were constructed at CERN within the framework of the ASACUSA collaboration of the Antiproton Decelerator facility. The DRS4 ASIC was procured from PSI LTP. S. Ritt provided technical advice for the design of our electronics. The  $E \times B$  Wien filter was developed by the PSI beamline operations group. The quadrupole doublet magnet was donated by the CERN magnet group. Y. Murakami, K. Todoroki and H. Yamada provided vital assistance during some phases of the experiments.

**Author contributions** M.H. conceived the experimental method and developed the DPSS Nd:YAG and OPG-OPA laser systems, 140-channel photomultiplier array with active breeders and high-voltage bias generators, scintillators, data acquisition electronics of the particle detectors, the concentric pair of cylindrical target chambers with 500- $\mu\text{m}$ -thick aluminium walls, cryogenic and beam delivery optics, and field-programmable gate array software. A.S. designed the cryogenic target, including its thermal radiation shields, the compact housing and supports for the particle detectors and electronics, and the scanning scintillation counter used to measure the diameter of the  $\pi^-$  beam, and carried out Monte Carlo simulations of the experiment. M.H., the CERN cryogenics group and A.S. designed the Joule–Thomson cryocooler and cryogenic pumping line used to cool the target to superfluid temperature. H.A.-K. designed the mechanical support of the cryogenic target including linear translation stages. D.B. wrote the data acquisition software. All authors contributed to the assembly of the experiment and data taking. D.B., A.S. and M.H. independently wrote analysis programs. M.H. wrote the manuscript and all authors discussed the results and contributed to the editing.

**Competing interests** The authors declare no competing interests.

### Additional information

**Correspondence and requests for materials** should be addressed to M.H.

**Peer review information** Nature thanks Klaus Jungmann and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>.



# Detection of metastable electronic states by Penning trap mass spectrometry

<https://doi.org/10.1038/s41586-020-2221-0>

Received: 15 January 2020

Accepted: 11 March 2020

Published online: 6 May 2020

 Check for updates

R. X. Schüssler<sup>1✉</sup>, H. Bekker<sup>1,7</sup>, M. Braß<sup>2</sup>, H. Cakir<sup>1</sup>, J. R. Crespo López-Urrutia<sup>1</sup>, M. Door<sup>1</sup>, P. Filianin<sup>1</sup>, Z. Harman<sup>1</sup>, M. W. Haverkort<sup>2</sup>, W. J. Huang<sup>1</sup>, P. Indelicato<sup>3</sup>, C. H. Keitel<sup>1</sup>, C. M. König<sup>1</sup>, K. Kromer<sup>1</sup>, M. Müller<sup>1</sup>, Y. N. Novikov<sup>4,5</sup>, A. Rischka<sup>1,8</sup>, C. Schweiger<sup>1</sup>, S. Sturm<sup>1</sup>, S. Ulmer<sup>6</sup>, S. Eliseev<sup>1✉</sup> & K. Blaum<sup>1</sup>

State-of-the-art optical clocks<sup>1</sup> achieve precisions of  $10^{-18}$  or better using ensembles of atoms in optical lattices<sup>2,3</sup> or individual ions in radio-frequency traps<sup>4,5</sup>. Promising candidates for use in atomic clocks are highly charged ions<sup>6</sup> (HCIs) and nuclear transitions<sup>7</sup>, which are largely insensitive to external perturbations and reach wavelengths beyond the optical range<sup>8</sup> that are accessible to frequency combs<sup>9</sup>. However, insufficiently accurate atomic structure calculations hinder the identification of suitable transitions in HCIs. Here we report the observation of a long-lived metastable electronic state in an HCI by measuring the mass difference between the ground and excited states in rhenium, providing a non-destructive, direct determination of an electronic excitation energy. The result is in agreement with advanced calculations. We use the high-precision Penning trap mass spectrometer PENTATRAP to measure the cyclotron frequency ratio of the ground state to the metastable state of the ion with a precision of  $10^{-11}$ —an improvement by a factor of ten compared with previous measurements<sup>10,11</sup>. With a lifetime of about 130 days, the potential soft-X-ray frequency reference at  $4.96 \times 10^{16}$  hertz (corresponding to a transition energy of 202 electronvolts) has a linewidth of only  $5 \times 10^{-8}$  hertz and one of the highest electronic quality factors ( $10^{24}$ ) measured experimentally so far. The low uncertainty of our method will enable searches for further soft-X-ray clock transitions<sup>8,12</sup> in HCIs, which are required for precision studies of fundamental physics<sup>6</sup>.

Modern clocks and frequency standards range from ensembles of neutral particles trapped in optical lattice clocks<sup>5,13</sup> to individual, singly charged ions confined in Paul traps<sup>14</sup>. With a fractional frequency accuracy of  $10^{-18}$ , such clocks enable stringent tests of fundamental symmetries (for example, Lorentz invariance<sup>15</sup>), geodetic measurements<sup>16</sup> and searches for new physics<sup>17</sup> and dark matter<sup>18</sup>. The transitions used as frequency references typically have long lifetimes, of the order of seconds, yielding sub-hertz linewidths; a recent example is the  $\text{Yb}^+$  clock<sup>15,19</sup>. Optical clocks use variations of the Ramsey method, driving such forbidden transitions through stimulated absorption and emission much faster than the natural transition rate, but maintaining the very long coherence times needed for the interrogation schemes owing to their small spontaneous emission rates<sup>1</sup>. One promising species for use in the next generation of clocks is HCIs<sup>6</sup> because their compact size in comparison with atoms makes them less sensitive to external field fluctuations. Several transitions have been proposed to feature a high sensitivity to a variation of fundamental constants<sup>20,21</sup>.

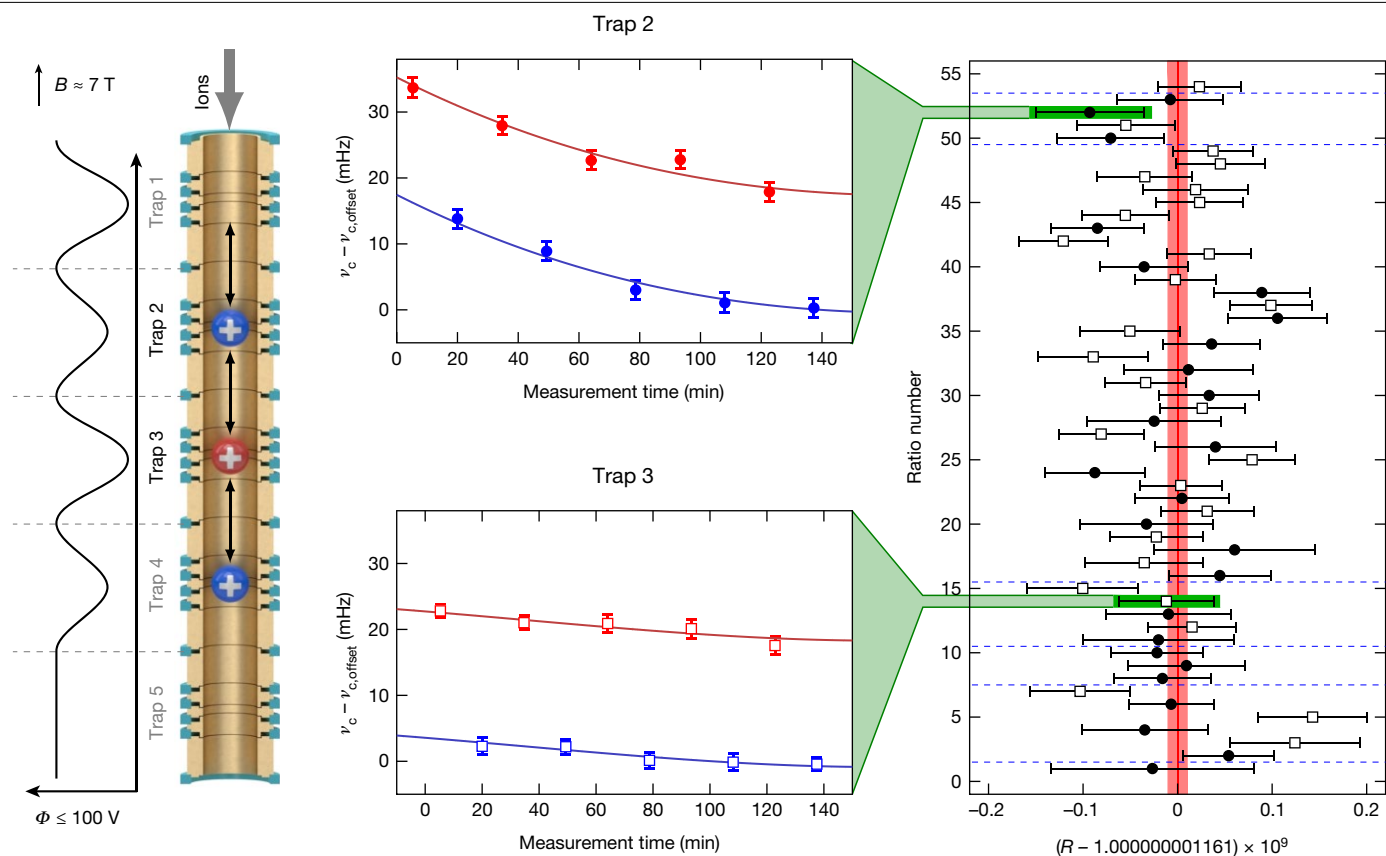
Although electronic binding energies in HCIs typically amount to several kiloelectronvolts<sup>22</sup> and inter-shell transitions usually appear in the X-ray region, there are also intra-shell fine and hyperfine transitions in the optical and ultraviolet range<sup>23–27</sup>. Furthermore, some HCIs

feature level crossings<sup>6,28</sup> with associated optical transitions. Most of these transitions are electric dipole (E1)-forbidden, and many are of the magnetic dipole (M1) type, with lifetimes of the order of milliseconds, but some are highly forbidden and can have extremely long lifetimes of up to millions of years<sup>17</sup>, as predicted theoretically. A few of these transitions with lifetimes ranging from milliseconds to seconds have been found and investigated using storage rings<sup>29</sup> and ion traps<sup>30,31</sup>, but so far no method has enabled a direct observation of the excitation energy of extremely long-lived metastable states.

It is difficult to theoretically predict such electronic energies in HCIs. Calculations suffer from difficulties related to strong relativistic contributions and complex correlations of several active electrons in open shells. Moreover, experimental identification of such transitions is challenging because their long lifetimes leads to sub-hertz linewidths and therefore requires very precise application of the Ritz–Rydberg method<sup>32</sup> to other, accurately known transitions to establish the searching range to be explored with narrow-linewidth spectroscopy lasers. Additionally, rough calculations with the flexible atomic code<sup>33</sup> can be used to estimate the energy of transitions.

Here we demonstrate how high-precision Penning trap mass spectrometry directly identifies a suitable clock transition in HCIs by

<sup>1</sup>Max Planck Institute for Nuclear Physics, Heidelberg, Germany. <sup>2</sup>Institute for Theoretical Physics, Heidelberg University, Heidelberg, Germany. <sup>3</sup>Laboratoire Kastler Brossel, Sorbonne Université, CNRS, ENS-PSL Research University, Collège de France, Paris, France. <sup>4</sup>Petersburg Nuclear Physics Institute, Gatchina, Russia. <sup>5</sup>St Petersburg State University, St Petersburg, Russia. <sup>6</sup>RIKEN, Fundamental Symmetries Laboratory, Wako, Japan. <sup>7</sup>Present address: Department of Physics, Columbia University, New York, NY, USA. <sup>8</sup>Present address: ARC Centre for Engineered Quantum Systems, School of Physics, The University of Sydney, Sydney, New South Wales, Australia. ✉e-mail: rima.schuessler@mpi-hd.mpg.de; sergey.eliseev@mpi-hd.mpg.de



**Fig. 1 | Results of the Re measurements at PENTATRAP.** All error bars indicate  $1\sigma$  uncertainty; for details see text. Left, three  $^{187}\text{Re}^{29+}$  ions (ion 1, blue; ion 2, red; ion 3, blue) are loaded from the top into the trap stack. For each measurement trap, the potential  $\Phi$  of the neighbouring traps is nominally identical; that is, the potential of trap 2 is also applied in trap 4. The magnetic field points along the direction of the trap axis. Centre, by shifting the ions one trap down or up after a cyclotron frequency determination, we obtain measurements with ion 1

(blue) and ion 2 (red) in trap 2 and with ion 2 (red) and ion 3 (second blue ion) in trap 3. Here, the ion 2 is in the ground state (red) and the other two (blue) are in the metastable electronic state. Right, all ratios ( $n = 54$ ) determined over seven measurement campaigns (divided by the dashed lines) display the stability of the system. The results for trap 2 (filled symbols) and trap 3 (empty symbols) show a similar behaviour. The final weighted mean is shown in red.

determining the mass difference of a  $^{187}\text{Re}^{29+}$  ion in the ground state and a metastable state to a precision below 2 eV or  $\sim 500$  THz. A relative mass measurement of the two HCl states with an unprecedented precision of  $1 \times 10^{-11}$  is achieved by determining the cyclotron frequency  $\nu_c = qB/(2\pi m)$  of the HCl with mass  $m$  and charge  $q$  in a strong magnetic field ( $B \approx 7$  T) and a weak electrostatic harmonic potential of the Penning trap. The cyclotron frequency is derived by measuring the ion trap frequencies (modified cyclotron frequency,  $\nu_+ \approx 16$  MHz; axial frequency,  $\nu_z \approx 700$  kHz; magnetron frequency,  $\nu_- \approx 10$  kHz) and applying the invariance theorem<sup>34</sup>  $\nu_c^2 = \nu_+^2 + \nu_z^2 + \nu_-^2$ .

### Penning trap mass spectrometer PENTATRAP

This work was carried out using the high-precision Penning trap mass spectrometer PENTATRAP at the Max-Planck-Institut für Kernphysik in Heidelberg. HCl<sup>+</sup> produced in the commercial electron beam ion trap DreBIT<sup>35</sup> using the MIVOC technique<sup>36</sup> are extracted in 1- $\mu\text{s}$ -long bunches with a kinetic energy of a few kiloelectronvolts per charge, selected according to their charge-to-mass ratio,  $q/m$ , by means of a  $90^\circ$  dipole magnet, and sent into the Penning traps. Prior to the trapping of the ions, their kinetic energy is reduced to a few electronvolts per charge by appropriately timed pulsed potentials applied to two cylindrical drift tubes. The 4-K cold bore of the superconducting magnet houses five cylindrical Penning traps<sup>37,38</sup>. Two of them (traps 2 and 3; see Fig. 1) are used to measure the trap frequencies of the ions of interest. Traps 1 and 4 serve to store ions, whereas trap 5 will allow monitoring

fluctuations of the trap potentials and the magnetic field in the future. To reduce temporal variations of the magnetic field in the traps, the temperature in the laboratory is kept constant at a level of 0.1 K per day. Furthermore, the level of liquid helium and the helium pressure inside the magnet cold bore are stabilized at a level of a fraction of a millimetre and 20  $\mu\text{bar}$ , respectively. These measures reduce fractional changes of the magnetic field in the measurement traps to values of less than a few parts in  $10^{10}$  per hour, resulting in a cyclotron frequency drift of only a few millihertz per hour at  $\nu_c = 16$  MHz. Because the magnet is actively shielded with a shielding factor of 50, the fractional fluctuation of the magnetic field in the trap due to the fluctuation of the stray magnetic field does not exceed 3 parts in  $10^{11}$ .

The measurement procedure of the cyclotron frequency ratio of two different ion states consists of the preparatory and the main phase. During the preparatory phase, three ions in the two states are first loaded into the three innermost traps (ion 1 in trap 2 and ion 3 in trap 4 being in the same state—for example, the ground state—whereas ion 2 in trap 3 being in the other state). The desired ion states are confirmed by a rough determination of their respective cyclotron frequencies; if they are not in the desired state, ions are reloaded. After that, their trap motion amplitudes are reduced (the ions are ‘cooled’) by coupling the magnetron and cyclotron motions to the axial motion, which in turn is brought into thermal equilibrium with a tank circuit having a temperature of around 6 K. The resonance frequency of the tank circuit is approximately equal to that of the axial motion. The axial, magnetron and cyclotron amplitudes of cooled  $^{187}\text{Re}^{29+}$  ions are approximately



10  $\mu\text{m}$ , 2  $\mu\text{m}$  and 2  $\mu\text{m}$ , respectively. Great care is taken to prepare only a single ion of interest in each trap, because the presence of another undesired ion in the trap would disturb the motion of the ion of interest and hence alter its motional frequencies. The main phase is devoted to the measurement of the ion trap frequencies in traps 2 and 3. The reduced cyclotron and axial frequencies are simultaneously measured by means of the pulse-and-phase<sup>39</sup> and dip<sup>40</sup> techniques, respectively. This minimizes systematic shifts arising from changes in the external fields compared with sequential measurements of the trap frequencies. The measurement cycle consists of a set of sub-measurements. First, the cyclotron motion is excited to an amplitude of approximately 10  $\mu\text{m}$  with a subsequent measurement of its phase (reference phase). It is followed by the re-cooling of all three motions. After that, the cyclotron motion is again excited to the same amplitude and is allowed to freely evolve its phase for 40 s during the subsequent measurement (measurement phase). During the phase evolution time of 40 s, the axial frequency is measured with the dip method. The measurement cycle is ended by re-cooling all ion motions. Because the relative mass difference between the ground and metastable states is at the level of  $10^{-9}$ , it is sufficient to measure the smallest trap frequency—the magnetron frequency—once a day with a moderate precision for the determination of the cyclotron frequency ratio. Ten measurement cycles constitute a measurement run. A unique feature of PENTATRAP is that during each run, frequency measurements are performed simultaneously in synchronized traps—for instance, on ion 1 in trap 2 (upper blue ion in Fig. 1) and ion 2 (red) in trap 3—for approximately 12 min. After that, the ion species are swapped between traps 2 and 3 by moving the three ions adiabatically one trap up. Then, a measurement run is carried out with ion 2 in trap 2 and ion 3 (lower blue ion) in trap 3. Lastly, the ion species are swapped back by moving the three ions one trap down again. After each change of the configuration of ions in the traps, the ion motions are cooled for 20 s. In this way, one alternately determines the cyclotron frequencies of ions in the ground and metastable states in each measurement trap. This measurement sequence is repeated until the measurement is stopped, given that—even though the vacuum allows storage times of several hours to days—ions are then lost owing to charge-exchange reactions with background gas. If determination of the metastable state is no longer possible, the traps are emptied and a new set of ions is loaded.

### Energy of the metastable state

To obtain the cyclotron frequency ratio of the ions in the ground and metastable states, the polynomial method described, for instance, in refs.<sup>10,11,41</sup> is applied in the following way. We treat the two measurement traps independently. The temporal variation of the magnetic field—and hence the cyclotron frequency—is approximated with a low-order polynomial, as shown in the middle plots of Fig. 1. This assumption is based on our experience with the magnetic field evolution over a scale of a few hours. The full set of measured cyclotron frequencies is divided into groups. The group length must be as short as possible, such that it can be described by a low-order polynomial, but must contain substantially more data points than the polynomial degrees of freedom. Thus, each group is chosen to be approximately 2 h long and to contain ten cyclotron frequency points (five points correspond to the ion in the ground state and five to the ion in the metastable state); see middle plots of Fig. 1. We use third-grade polynomials because this is the lowest order that has at least one inflection point. Data points from ions in both the metastable and the ground state are fitted with the same polynomial using global fitting parameters. The polynomial for the data points from the ion in the ground state is scaled with an additional fitting parameter  $R$ , which corresponds to the frequency ratio. This yields three quantities: (1) the frequency ratio  $R$ , (2) its uncertainty  $\Delta R$  and (3) the reduced  $\chi^2$ . If  $\chi^2 > 1$ , we scale the uncertainty of the ratio by  $\chi^2$ . The final ratio  $R$  is obtained as a weighted mean of the frequency

ratios from all groups and measurement traps. Its statistical uncertainty  $\Delta R$  is the larger of the internal and external errors<sup>42</sup>; see right side of Fig. 1. A variation of the polynomial order (second, third and fourth) results in an increase of the statistical uncertainty of the final ratio by  $\sim 10\%$ .

The data were blindly analysed independently by two people, each using separately developed software. Both analyses yield similar external errors of  $8 \times 10^{-12}$  and internal errors of  $7 \times 10^{-12}$ , resulting in a Birge ratio of 1.14. The two weighted means differ by  $4 \times 10^{-12}$ . Thus, the combined final ratio is

$$R - 1 = 1.161(10)_{\text{stat}} \times 10^{-9}$$

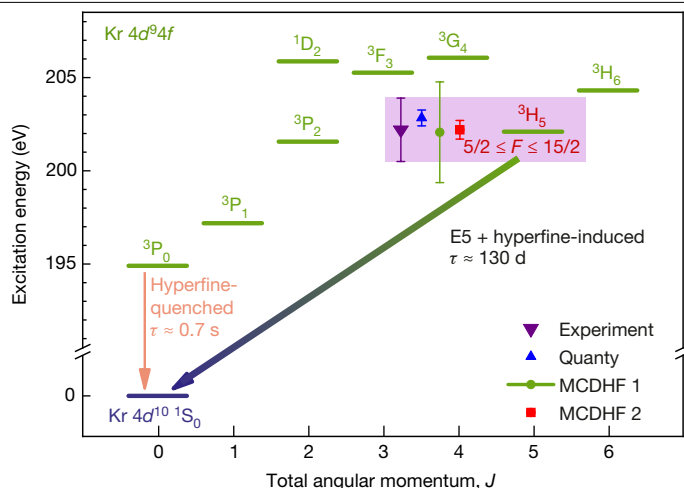
(All reported uncertainties are  $1\sigma$ .) The uniqueness of the considered  $^{187}\text{Re}^{29+}$  ions in the two electron configuration states is that the fractional difference in their masses is only  $10^{-9}$ . Uncertainties due to systematic effects such as image charge shift, relativistic shift and higher-order magnetic and electric field (see, for example, ref.<sup>43</sup>) variations can be neglected. For these uncertainties to be at the level of 10% of the statistical uncertainty ( $10^{-12}$ ), the cyclotron frequencies would have to be shifted by at least 10 kHz, which is excluded considering the present known performance of PENTATRAP. Therefore, the systematic uncertainties in the final ratio are well below the statistical uncertainty and hence can be neglected.

The ground and metastable states in  $^{187}\text{Re}^{29+}$  belong to the  $\text{Kr } 4d^{10}1s_0$  and  $4d^9 4f^3 3H_5$  electron configurations, respectively. Because the metastable state can only decay to the ground state—by a highly forbidden electric triacontadipole (E5) transition—its estimated lifetime is  $\sim 130$  d. (We note that whereas optical transitions with similarly high multipolarities may feature much longer lifetimes<sup>6</sup>, the high transition energy, together with hyperfine quenching, yields a lifetime in this regime.) Owing to repeated and fast processes of electron impact excitation followed by radiative decay inside the electron beam ion trap, low-lying metastable states and the ground state are roughly equally populated. In this case, the metastable ion fraction detected in PENTATRAP is approximately 50%. For the determination of the energy of the metastable state, seven measurement campaigns with three ions each were performed over the course of 12 days, during which not a single decay was detected.

From the final ratio, the energy of the metastable state with respect to the ground state can be calculated using  $\Delta E_{\text{Re}} = m(^{187}\text{Re}^{29+})(R - 1)c^2$  ( $c$ , speed of light in vacuum). The uncertainty of the ion mass enters the total uncertainty reduced by a factor of  $(R - 1)$ , and must be known to only  $\Delta m/m \approx 10^{-4}$  to become negligible. Because the mass of the 29 removed electrons and their binding energies are well below that limit, we can simply use the neutral mass of  $^{187}\text{Re}$  for the calculation, which gives  $\Delta E_{\text{Re}} = 202.2(17)$  eV.

### Calculations of the energy difference

The high charge state and the complexity of the open  $4d$  and  $4f$  subshells calls for state-of-the-art theoretical methods to match the experimental accuracy. We apply three different fully relativistic approaches, namely, the multiconfiguration Dirac–Hartree–Fock (MCDHF) method<sup>44,45</sup> in two different implementations and the configuration interaction method employing Kohn–Sham orbitals, as implemented by the many-body script language Quanty<sup>46</sup>. The calculations also help to identify the observed metastable state. The first such state could be the  $4d^9 4f^3 P_0$  level with an energy of  $\sim 195$  eV (see Fig. 2). This level may decay only to the ground state by a highly forbidden two-photon transition. However,  $^{187}\text{Re}$  has a nuclear spin of  $I = 5/2$  and thus a magnetic moment that couples this  $^3P_0$  level to the nearby  $^3P_1$  level, which can decay to the ground state by an E1 transition. The  $^3P_0$  state thus acquires part of the probability of this allowed transition, a phenomenon called hyperfine quenching<sup>47–50</sup>. Following ref.<sup>47</sup> and using the



**Fig. 2 | The  $4d^{10}$  ground state and relevant  $4d^9 4f$  excited electronic states of the  $^{187}\text{Re}^{29+}$  ion.** Experimental result (shaded area) and theoretical values obtained using MCDHF approaches in two different implementations (MCDHF 1 and 2) and a configuration-interaction calculation (Quanta). All error bars indicate  $1\sigma$  uncertainty. The level structure is taken from the MCDHF 1 calculations. The lifetimes  $\tau$  of the levels of interest are denoted and the main decay processes (hyperfine-induced/quenched and E5) are stated.

MCDHF code<sup>51</sup>, we estimate a lifetime of 0.7 s, which is too short to be observed. The observed state must then be the next metastable level,  $4d^9 4f^3 H_5$ , which—neglecting quenching—can decay to the ground state only by an E5 transition with a lifetime of ~239 d. Decays to the levels just below—that is, to the  $3P_0$ ,  $3P_1$  and  $3P_2$  states—have very low probabilities because of their high multiplicities ( $M5$ ,  $E4$  and  $M3$ , respectively) and small photon energies. The  $3H_5$  state is also subject to hyperfine quenching; for example, its lifetime with a coupled atomic and nuclear angular momentum of  $F=7/2$  reduces to  $\sim 124$  d. This corresponds to a width-to-transition energy ratio of  $3 \times 10^{-25}$ , enabling very-high-precision measurements. The states with  $F=9/2$  and  $F=11/2$  live 128 d, and other states coupled to more distant fine-structure levels have even longer lifetimes.

In the Quanta calculations of the transition energies, the relativistic Dirac Hamiltonian includes the Coulomb and Breit interactions between all electrons. The Fock space is spanned by multi-Slater-determinant states constructed from relativistic Kohn–Sham single-electron states, obtained from the density functional code FPLO<sup>52</sup>, using the local spin density functional of ref.<sup>53</sup>. Within this calculation we explicitly include shells with principal quantum numbers  $n < n_{\text{max}} = 7$ . To extrapolate to  $n_{\text{max}} = \infty$ , we compare results for calculations with  $n < 7$ ,  $n < 6$  and  $n < 5$ . Allowing quantum fluctuations within the  $n=4$  shells, that is, including configurations with holes in the  $4s$ – $4d$  shells and additional electrons in the  $4f$  shell, we calculate the energies of the  $4d^9 4f$  excited states in a first approximation. Allowing for a maximum of three electrons in the  $4f$  shell suffices to calculate the excitation energy with the required accuracy. Next, we extend the configuration space to allow electrons to scatter between the  $n=4$ , 5 and 6 shells, where at most three holes are excited. By successively adding further configurations, we can extrapolate the excitation energy of the  $3H_5$  metastable state to be between 202.41 eV and 203.26 eV.

In the MCDHF calculations, we optimize not only the mixing coefficients of the configurations but also the radial single-electron wave functions by self-consistently solving the corresponding radial Dirac equations. Quantum electrodynamics corrections to the excitation energy are estimated by the model potential method<sup>54</sup> and by computing the self-energy shift of the  $4d$  and  $4f$  valence electrons using effective potentials, which account for screening by the core electrons. One of the MCDHF calculations (MCDHF1) employs MCDHF code<sup>51</sup>, including

full relaxation of all spectroscopic orbitals, as well as single and double excitations from all  $n=4$  orbitals to the free single-electron states up to  $5d$ , with a result of 202.1(27) eV. In the other computation (MCDHF2), using GRASP2018<sup>55</sup>, we generate the set of configurations with single and double excitations from the  $n=4$  states up to  $9h$ , with the virtual orbitals optimized layer by layer. Subsequently, the effect of triple electron exchanges is accounted for in a configuration interaction approach. The convergence of the results is monitored while systematically expanding the configuration set, allowing us to estimate the uncertainty of the calculations and providing an extrapolated value of 202.2(5) eV.

## Conclusion

Figure 2 shows a comparison of the three theoretical values with the experimental one. As a crosscheck, we also performed frequency ratio measurements (with less statistics) of the same metastable state in  $\text{Os}^{30+}$ , which is isoelectronic to  $\text{Re}^{29+}$ . We obtained  $\Delta E_{\text{Os}} = 207(3)$  eV. As the electronic binding energy roughly scales with  $Z^2$  ( $Z$ , atomic number), this higher energy is explained by the additional proton in Os ( $Z = 76$ ) and the experimental result is again in agreement with the theoretical predictions, which are not explicitly given.

In summary, we have demonstrated a method of determining the excitation energies of extremely long-lived metastable states in HCLs that yields realistic measurement times if the metastable fraction in the atomic or ionic sample is at least 10% and the lifetime is longer than a few hours. In principle, with the precision achieved here for Re ions, it would also be possible to detect nuclear transitions with energies as low as 5 eV if the abundance of the isomeric nuclear state in the ions is comparable to that of the ground state.

## Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-020-2221-0>.

- Ludlow, A. D. et al. Optical atomic clocks. *Rev. Mod. Phys.* **87**, 637–701 (2015).
- Katori, H. Optical lattice clocks and quantum metrology. *Nat. Photon.* **5**, 203–210 (2011).
- Marti, G. E. et al. Imaging optical frequencies with 100  $\mu\text{Hz}$  precision and 1.1  $\mu\text{m}$  resolution. *Phys. Rev. Lett.* **120**, 103201 (2018).
- Brewer, S. M. et al. An  $^{27}\text{Al}^+$  quantum-logic clock with systematic uncertainty below  $10^{-18}$ . *Phys. Rev. Lett.* **123**, 033201 (2019).
- Nicholson, T. et al. Systematic evaluation of an atomic clock at  $2 \times 10^{-18}$  total uncertainty. *Nat. Commun.* **6**, 6896 (2015).
- Kozlov, M. G., Safronova, M. S., Crespo López-Urrutia, J. R. & Schmidt, P. O. Highly charged ions: optical clocks and applications in fundamental physics. *Rev. Mod. Phys.* **90**, 045005 (2018).
- Seiferle, B. et al. Energy of the  $^{229}\text{Th}$  nuclear clock transition. *Nature* **573**, 243–246 (2019).
- Nauta, J. et al. Towards precision measurements on highly charged ions using a high harmonic generation frequency comb. *Nucl. Instrum. Methods Phys. Res. B* **408**, 285–288 (2017).
- Cingöz, A. et al. Direct frequency comb spectroscopy in the extreme ultraviolet. *Nature* **482**, 68–71 (2012).
- Shi, W., Redshaw, M. & Myers, E. G. Atomic masses of  $^{32,33}\text{S}$ ,  $^{84,86}\text{Kr}$ , and  $^{129,132}\text{Xe}$  with uncertainties  $\leq 1$  ppb. *Phys. Rev. A* **72**, 022510 (2005).
- Eliseev, S. et al. Direct measurement of the mass difference of  $^{163}\text{Ho}$  and  $^{163}\text{Dy}$  solves the Q-value puzzle for the neutrino mass determination. *Phys. Rev. Lett.* **115**, 062501 (2015).
- Crespo López-Urrutia, J. R. Frequency metrology using highly charged ions. *J. Phys. Conf. Ser.* **723**, 012052 (2016).
- Ushijima, I. et al. Cryogenic optical lattice clocks. *Nat. Photon.* **9**, 185–189 (2015).
- Huntemann, N. et al. Single-ion atomic clock with  $3 \times 10^{-18}$  systematic uncertainty. *Phys. Rev. Lett.* **116**, 063001 (2016).
- Sanner, C. et al. Optical clock comparison for Lorentz symmetry testing. *Nature* **567**, 204–208 (2019).
- Mehlstäubler, T. E. et al. Atomic clocks for geodesy. *Rep. Prog. Phys.* **81**, 064401 (2018).
- Safronova, M. et al. Search for new physics with atoms and molecules. *Rev. Mod. Phys.* **90**, 025008 (2018).
- Derevianko, A. & Pospelov, M. Hunting for topological dark matter with atomic clocks. *Nat. Phys.* **10**, 933–936 (2014).
- McGrew, W. F. et al. Atomic clock performance enabling geodesy below the centimetre level. *Nature* **564**, 87–90 (2018).

20. Ong, A., Berengut, J. C. & Flambaum, V. V. in *Fundamental Physics in Particle Traps* 293 (Springer, 2014).
21. Dzuba, V. A. & Flambaum, V. V. Highly charged ions for atomic clocks and search for variation of the fine structure constant. *Hyperfine Interact.* **236**, 79–86 (2015).
22. Gillaspay, J. D. Highly charged ions. *J. Phys. B* **34**, R39 (2001).
23. Klaft, I. et al. Precision laser spectroscopy of the ground state hyperfine splitting of hydrogenlike  $^{209}\text{Bi}^{82+}$ . *Phys. Rev. Lett.* **73**, 2425–2427 (1994).
24. Morgan, C. A. et al. Observation of visible and UV magnetic dipole transitions in highly charged xenon and barium. *Phys. Rev. Lett.* **74**, 1716–1719 (1995).
25. Crespo López-Urrutia, J. R., Beiersdorfer, P., Savin, D. & Widmann, K. Direct observation of the spontaneous emission of the hyperfine transition  $F = 4$  to  $F = 3$  in ground state hydrogenlike  $^{165}\text{Ho}^{66+}$  in an electron beam ion trap. *Phys. Rev. Lett.* **77**, 826–829 (1996).
26. Schiller, S. Hydrogenlike highly charged ions for tests of the time independence of fundamental constants. *Phys. Rev. Lett.* **98**, 180801 (2007).
27. Crespo López-Urrutia, J. R. The visible spectrum of highly charged ions: a window to fundamental physics. *Can. J. Phys.* **86**, 111–123 (2008).
28. Berengut, J. C., Dzuba, V. A., Flambaum, V. V. & Ong, A. Highly charged ions with E1, M1, and E2 transitions within laser range. *Phys. Rev. A* **86**, 022517 (2012).
29. Schippers, S. et al. Storage-ring measurement of the hyperfine induced  $^{47}\text{Tl}^{80}(2s2p\ ^3P_0 \rightarrow 2s^2\ ^1S_0)$  transition rate. *Phys. Rev. Lett.* **98**, 033001 (2007).
30. Träbert, E. et al. Time-resolved soft-X-ray spectroscopy of a magnetic octupole transition in nickel-like xenon, cesium, and barium ions. *Phys. Rev. A* **73**, 022508 (2006).
31. Träbert, E., Beiersdorfer, P. & Brown, G. V. Observation of hyperfine mixing in measurements of a magnetic octupole decay in isotopically pure nickel-like  $^{128}\text{Xe}$  and  $^{132}\text{Xe}$  ions. *Phys. Rev. Lett.* **98**, 263001 (2007).
32. Ritz, W. On a new law of series spectra. *Astrophys. J.* **28**, 237–243 (1908).
33. Gu, M. F. The flexible atomic code. *Can. J. Phys.* **86**, 675–689 (2008).
34. Brown, L. S. & Gabrielse, G. Geonium theory: physics of a single electron or ion in a Penning trap. *Rev. Mod. Phys.* **58**, 233–311 (1986).
35. Zschornak, G. et al. *Dresden Electron Beam Ion Sources: Latest Developments*. Technical Report (Technische Universität Dresden & DREBIT GmbH, 2009); [https://www.researchgate.net/publication/228896380\\_DRESDEN\\_ELECTRON\\_BEAM\\_ION\\_SOURCES\\_LATEST\\_DEVELOPMENTS](https://www.researchgate.net/publication/228896380_DRESDEN_ELECTRON_BEAM_ION_SOURCES_LATEST_DEVELOPMENTS).
36. Koivisto, H., Arje, J. & Nurmia, M. Metal ions from the volatile compounds method for the production of metal ion beams. *Rev. Sci. Instrum.* **69**, 785–787 (1998).
37. Roux, C. et al. The trap design of PENTATRAP. *Appl. Phys. B* **107**, 997–1005 (2012).
38. Repp, J. et al. PENTATRAP: a novel cryogenic multi-Penning-trap experiment for high-precision mass measurements on highly charged ions. *Appl. Phys. B* **107**, 983–996 (2012).
39. Cornell, E. A. et al. Single-ion cyclotron resonance measurement of  $M(\text{CO}^+)/M(\text{N}_2^+)$ . *Phys. Rev. Lett.* **63**, 1674–1677 (1989).
40. Feng, X. et al. Tank circuit model applied to particles in a Penning trap. *J. Appl. Phys.* **79**, 8–13 (1996).
41. Karthein, J. et al. QEC-value determination for  $^{21}\text{Na} \rightarrow ^{21}\text{Ne}$  and  $^{23}\text{Mg} \rightarrow ^{23}\text{Na}$  mirror-nuclei decays using high-precision mass spectrometry with ISOLTRAP at the CERN ISOLDE facility. *Phys. Rev. C* **100**, 015502 (2019).
42. Birge, R. T. The calculation of errors by the method of least squares. *Phys. Rev.* **40**, 207–227 (1932).
43. Smorra, C. et al. A parts-per-billion measurement of the antiproton magnetic moment. *Nature* **550**, 371–374 (2017).
44. Grant, I. P. Relativistic calculation of atomic structures. *Adv. Phys.* **19**, 747–811 (1970).
45. Desclaux, J. P., Mayers, D. F. & O'Brien, F. Relativistic atomic wave functions. *J. Phys. B* **4**, 631–642 (1971).
46. Haverkort, M. W. Quanta for core level spectroscopy – excitons, resonances and band excitations in time and frequency domain. *J. Phys. Conf. Ser.* **712**, 012001 (2016).
47. Indelicato, P., Parente, F. & Marrus, R. Effect of hyperfine structure on the  $2^3P_0$  and the  $2^3P_1$  lifetime in heliumlike ions. *Phys. Rev. A* **40**, 3505–3514 (1989).
48. Gould, H., Marrus, R. & Mohr, P. Radiative decay of the  $2^3S_1$  and  $2^3P_2$  states of heliumlike vanadium ( $Z = 23$ ) and iron ( $Z = 26$ ). *Phys. Rev. Lett.* **33**, 676–680 (1974).
49. Indelicato, P. et al. Hyperfine quenching and precision measurement of the  $2^3P_0 - 2^3P_1$  fine structure splitting in heliumlike gadolinium ( $\text{Gd}^{62+}$ ). *Phys. Rev. Lett.* **68**, 1307–1310 (1992).
50. Birkett, B. et al. Hyperfine quenching and measurement of the  $2^3P_0 - 2^3P_1$  fine-structure splitting in heliumlike silver ( $\text{Ag}^{45+}$ ). *Phys. Rev. A* **47**, R2454–R2457 (1993).
51. Indelicato, P. & Desclaux, J. MCDFGME: a multiconfiguration Dirac-Fock and general matrix elements program (2005).
52. Koepf, K. & Eschrig, H. Full-potential nonorthogonal local-orbital minimum-basis band-structure scheme. *Phys. Rev. B* **59**, 1743–1757 (1999).
53. Perdew, J. P. & Wang, Y. Accurate and simple analytic representation of the electron-gas correlation energy. *Phys. Rev. B* **45**, 13244–13249 (1992).
54. Shabaev, V. M., Tupitsyn, I. I. & Yerokhin, V. A. QED-MOD: Fortran program for calculating the model Lamb-shift operator. *Comput. Phys. Commun.* **189**, 175–181 (2015).
55. Froese Fischer, C., Gaigalas, G., Jönsson, P. & Bierón, J. GRASP2018 – a Fortran 95 version of the General Relativistic Atomic Structure Package. *Comput. Phys. Commun.* **237**, 184–187 (2019).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2020



## Data availability

The datasets analysed in this study are available from the corresponding author.

## Code availability

The experimental data were analysed using OriginLab and a self-written analysis script, which is available from the corresponding author. The Quanty code and its documentation are available from <http://www.quanty.org>. MCDHF 1 is described in ref.<sup>51</sup> and MCDHF 2 in ref.<sup>55</sup>.

**Acknowledgements** This article comprises parts of the PhD thesis work of H.C., to be submitted to Heidelberg University, Germany. This work is supported by the German Research Foundation (DFG) Collaborative Research Centre SFB 1225 (ISOQUANT) and by the DFG Research UNIT FOR 2202. P.I. acknowledges partial support from NIST. Laboratoire Kastler Brossel (LKB) is supported by Unité Mixte de Recherche de Sorbonne Université, de ENS-PSL Research University, du Collège de France et du CNRS No. 8552. P.I., Y.N.N. and K.B. are

members of the Allianz Program of the Helmholtz Association, contract number EMMI HA-216 ‘Extremes of Density and Temperature: Cosmic Matter in the Laboratory’. P.I. thanks J.-P. Desclaux for help in improving the MCDFGME code. This project received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme under grant agreement number 832848 - FunI. Furthermore, we acknowledge funding and support by the International Max Planck Research School for Precision Tests of Fundamental Symmetries (IMPRS-PTFS) and by the Max Planck, RIKEN, PTB Center for Time, Constants and Fundamental Symmetries.

**Author contributions** The experiment was performed by R.X.S., M.D., A.R. and S.E. The data were analysed by R.X.S. and S.E. Theoretical calculations were performed by H.B., M.B., H.C., Z.H., M.W.H. and P.I. The manuscript was written by R.X.S., M.B., Z.H., M.W.H., P.I., J.R.C.L.-U. and S.E. and edited by S.U. and K.B. All authors discussed and approved the data as well as the manuscript.

**Competing interests** The authors declare no competing interests.

### Additional information

**Correspondence and** requests for materials should be addressed to R.X.S. or S.E.

**Peer review information** *Nature* thanks Jens Dilling, Marianna Safronova and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>.

# Mapping the twist-angle disorder and Landau levels in magic-angle graphene

<https://doi.org/10.1038/s41586-020-2255-3>

Received: 25 July 2019

Accepted: 7 February 2020

Published online: 6 May 2020

 Check for updates

A. Uri<sup>1,9</sup>, S. Grover<sup>1,9</sup>, Y. Cao<sup>2,9</sup>, J. A. Crosse<sup>3,4</sup>, K. Bagani<sup>1</sup>, D. Rodan-Legrain<sup>2</sup>, Y. Myasoedov<sup>1</sup>, K. Watanabe<sup>5</sup>, T. Taniguchi<sup>5</sup>, P. Moon<sup>3,4,6,7</sup>, M. Koshino<sup>8</sup>, P. Jarillo-Herrero<sup>2,✉</sup> & E. Zeldov<sup>1,✉</sup>

The recently discovered flat electronic bands and strongly correlated and superconducting phases in magic-angle twisted bilayer graphene (MATBG)<sup>1,2</sup> crucially depend on the interlayer twist angle,  $\theta$ . Although control of the global  $\theta$  with a precision of about 0.1 degrees has been demonstrated<sup>1–7</sup>, little information is available on the distribution of the local twist angles. Here we use a nanoscale on-tip scanning superconducting quantum interference device (SQUID-on-tip)<sup>8</sup> to obtain tomographic images of the Landau levels in the quantum Hall state<sup>9</sup> and to map the local  $\theta$  variations in hexagonal boron nitride (hBN)-encapsulated MATBG devices with relative precision better than 0.002 degrees and a spatial resolution of a few moiré periods. We find a correlation between the degree of  $\theta$  disorder and the quality of the MATBG transport characteristics and show that even state-of-the-art devices—which exhibit correlated states, Landau fans and superconductivity—display considerable local variation in  $\theta$  of up to 0.1 degrees, exhibiting substantial gradients and networks of jumps, and may contain areas with no local MATBG behaviour. We observe that the correlated states in MATBG are particularly fragile with respect to the twist-angle disorder. We also show that the gradients of  $\theta$  generate large gate-tunable in-plane electric fields, unscreened even in the metallic regions, which profoundly alter the quantum Hall state by forming edge channels in the bulk of the sample and may affect the phase diagram of the correlated and superconducting states. We thus establish the importance of  $\theta$  disorder as an unconventional type of disorder enabling the use of twist-angle gradients for bandstructure engineering, for realization of correlated phenomena and for gate-tunable built-in planar electric fields for device applications.

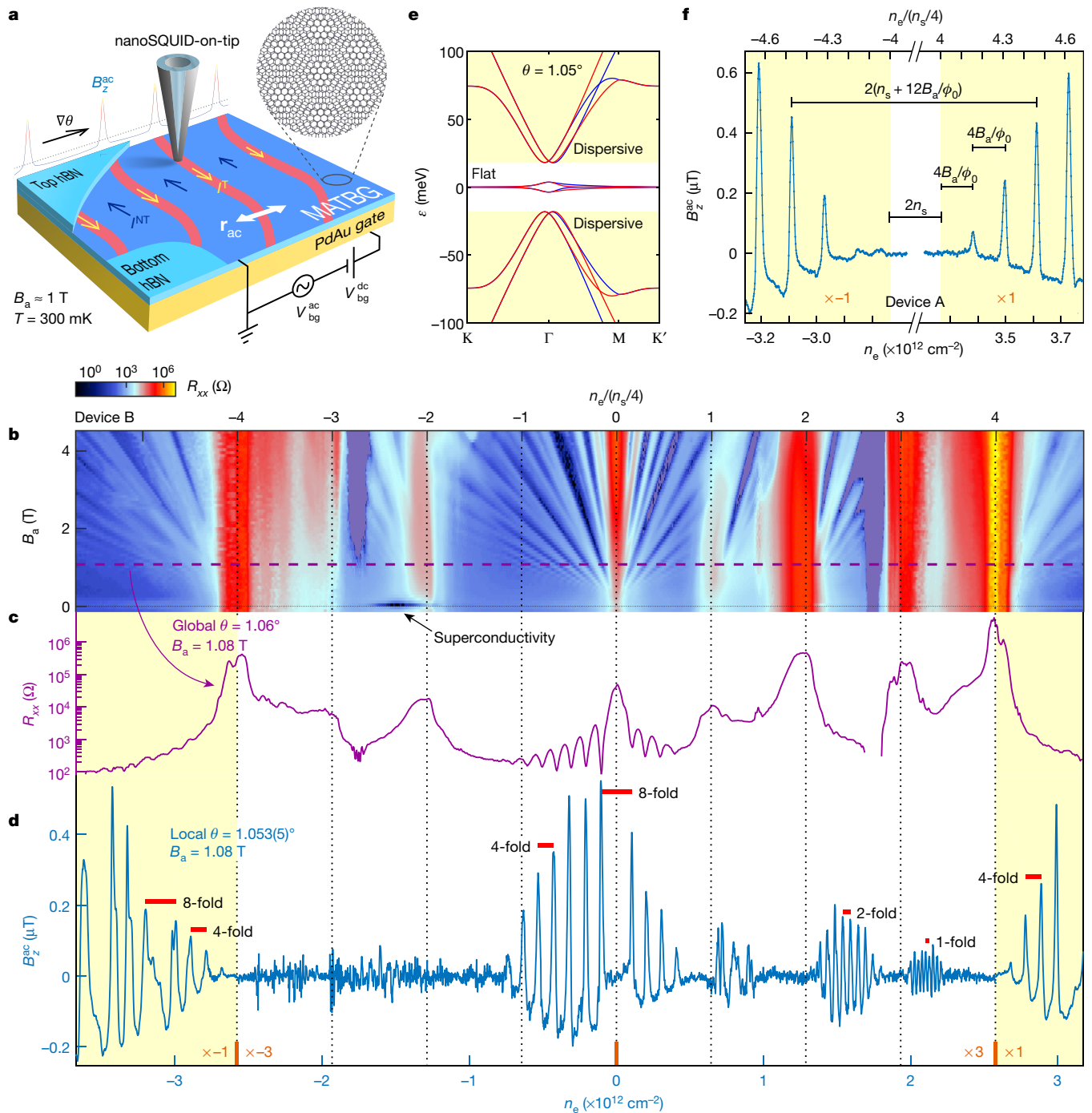
Strong electronic correlations arise in twisted bilayer graphene when the low energy bands become exceedingly narrow in the vicinity of the magic angle<sup>1–7</sup>,  $\theta_M \approx 1.1^\circ$ . Initial estimates of the bandwidth of these flat bands assumed a rigid and uniform rotation between the two graphene sheets, leading to a moiré pattern<sup>10–13</sup>. Recent bandstructure calculations have shown, however, that twist-angle relaxation within a single supercell (about 13 nm for  $\theta \approx 1.1^\circ$ ), results in electronic reconstruction that substantially modifies the bandstructure<sup>14,15</sup>. Because the bandstructure of the flat bands is determined on a scale of several supercells, twist angle gradients—similarly to the predicted strong effects of heterostrain<sup>16,17</sup>—should modify the single-particle bandstructure and induce symmetry breaking, possibly leading to properties that have not yet been considered. Moreover, as correlated phenomena may occur owing to electronic interactions on distances larger than the supercell, twist-angle variations may affect the stability of the competing orders, enriching the phase diagram of the correlated states.

Scanning tunnelling microscopy studies have shown that the local twist angle can vary substantially in the same sample; these studies have also observed stacking faults and structural defects<sup>18–24</sup>. Large

inhomogeneities and extensive networks of stacking faults in bilayer graphene have also been observed by transmission electron microscopy<sup>15,25–27</sup>. In this work, we provide high-resolution maps of  $\theta(\mathbf{r})$  in hBN-encapsulated MATBG devices (here  $\mathbf{r} = (x, y)$  is the position within the sample). The results reveal sizeable twist-angle gradients that constitute an unconventional type of disorder that strongly affects both the stability of correlated phases and magneto-transport characteristics in twisted bilayer graphene.

We present here local studies of two samples (devices A and B, see Methods), fabricated using the ‘tear-and-stack’ technique<sup>28,29</sup>. Their global transport data show characteristic MATBG features<sup>1–7</sup> including superconductivity, correlated insulator states at integer fractions of  $n_s$  (four electrons per moiré supercell), and Landau fans, from which a global twist angle  $\theta = 1.06^\circ$  is derived (device B, Fig. 1b). We derive local maps of  $\theta(\mathbf{r})$  by imaging the structure of the Landau levels throughout the sample. In a conventional quantum Hall state, alternating compressible and incompressible strips are formed near the sample edges, where the Fermi energy  $\epsilon_F$  correspondingly resides within the Landau levels or in the energy gaps between them. These

<sup>1</sup>Department of Condensed Matter Physics, Weizmann Institute of Science, Rehovot, Israel. <sup>2</sup>Department of Physics, Massachusetts Institute of Technology, Cambridge, MA, USA. <sup>3</sup>Division of Arts and Sciences, NYU Shanghai, Shanghai, China. <sup>4</sup>NYU-ECNU Institute of Physics at NYU Shanghai, Shanghai, China. <sup>5</sup>National Institute for Material Science, Tsukuba, Japan. <sup>6</sup>Department of Physics, New York University, New York, NY, USA. <sup>7</sup>State Key Laboratory of Precision Spectroscopy, East China Normal University, Shanghai, China. <sup>8</sup>Department of Physics, Osaka University, Toyonaka, Japan. <sup>9</sup>These authors contributed equally: A. Uri, S. Grover, Y. Cao. ✉e-mail: [pjarillo@mit.edu](mailto:pjarillo@mit.edu); [eli.zeldov@weizmann.ac.il](mailto:eli.zeldov@weizmann.ac.il)



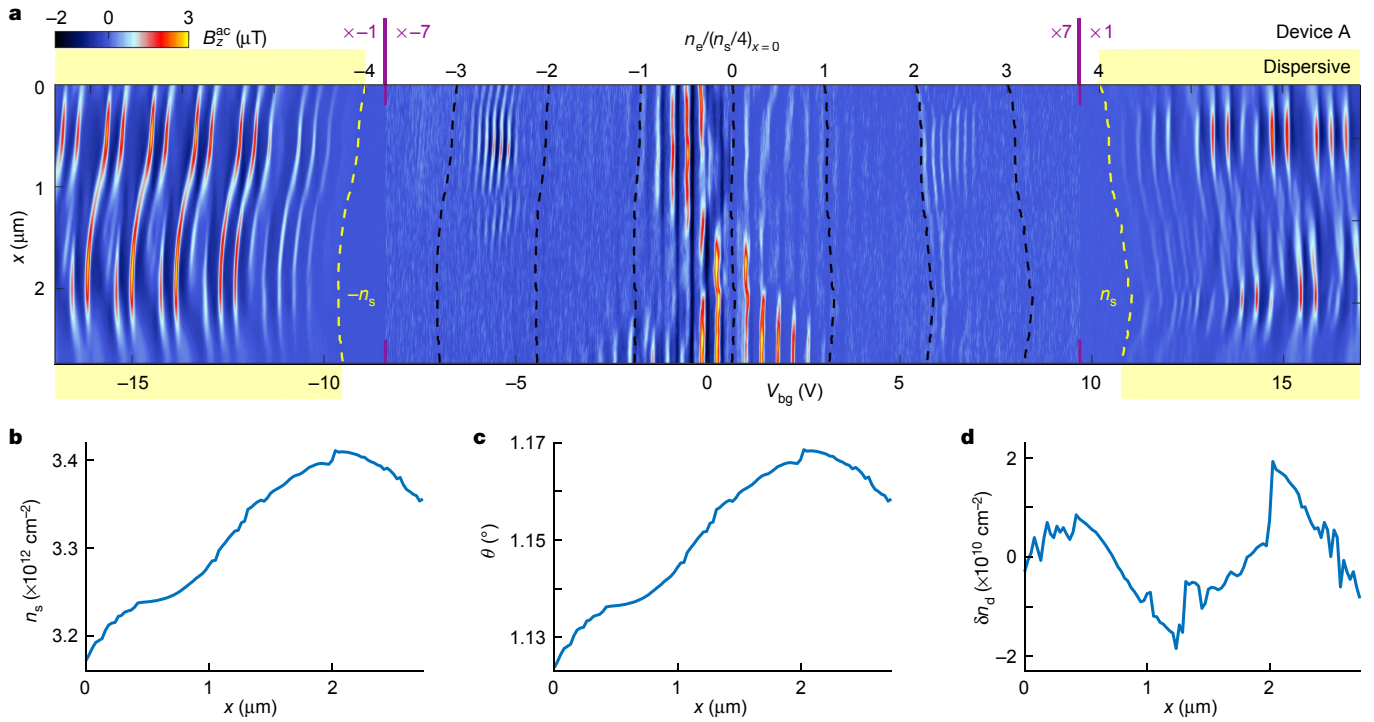
**Fig. 1 | Comparison between the global and local quantum Hall signatures in flat and dispersive bands in MATBG.** **a**, Experimental setup schematics with SOT scanning over MATBG (blue and magnified view) encapsulated in hBN (light blue). A voltage  $V_{bg}^{dc} + V_{bg}^{ac}$  is applied between the PdAu backgate and the grounded MATBG. The twist-angle gradient  $\nabla\theta$  induces an internal electric field and counterpropagating equilibrium quantum Hall topological  $I^T$  and nontopological  $I^{NT}$  currents in narrow incompressible (red) and wider compressible (blue) strips, respectively, flowing along equi- $\theta$  contours and detected by  $B_z^{ac}$ .  $V_{bg}^{ac}$  causes the strips to oscillate in their position with amplitude  $r_{ac}$ . **b**, Global  $R_{xx}$  versus electron density  $n_e$  and  $B_a$  for device B, showing insulating states at integer fillings  $n_e/(n_s/4)$ , Landau fans and superconductivity. **c**,  $R_{xx}(n_e)$  at  $B_a = 1.08$  T (extracted from **b** along the dashed

strips respectively carry counterpropagating nontopological ( $I^{NT}$ ) and topological ( $I^T$ ) equilibrium currents, as demonstrated recently in graphene<sup>9</sup>. By contrast, in MATBG these strips are found in the bulk of the sample instead of along the edges (Fig. 1a).

purple line). A global value  $\theta = 1.06^\circ$  is determined from the transport data. **d**,  $B_z^{ac}$  measured at a point in the bulk of device B versus  $n_e$  at  $B_a = 1.08$  T. The sharp  $B_z^{ac}$  peaks reflect  $I^T$  in incompressible strips with the sign determined by  $\sigma_{xx}$ , the magnitude by the energy gap of the Landau level and the separation by the Landau level degeneracy (red bars). The dispersive bands are shaded in yellow, the signal in the flat bands is amplified three times, and the p-band signal is multiplied by  $-1$  for clarity. From the position of the  $B_z^{ac}$  peaks the local twist angle is determined to be  $\theta = 1.053 \pm 0.005^\circ$ . **e**, Calculated bandstructure with flat (white) and dispersive (yellow) bands indicated. Blue and red represent the two valleys. **f**, Magnified  $B_z^{ac}$  peaks in the dispersive bands for device A at  $B_a = 1.19$  T, illustrating the procedure for determining the local  $n_s$  and the corresponding local  $\theta$ . The p-band signal is multiplied by  $-1$ .

We image these currents using a superconducting quantum interference device fabricated on the apex of a sharp pipette<sup>8,9</sup> (SQUID-on-tip, SOT; Fig. 1a). The Pb SOT, with a typical diameter  $d \approx 200$  nm, is scanned at a height of  $h_{SOT} \approx 30$  nm above the sample surface at  $T = 300$  mK in





**Fig. 2 | Structure of the Landau levels and derivation of the twist angle along a line scan.** **a**,  $B_z^{\text{ac}}(x)$  versus  $V_{\text{bg}}$  for device A acquired along the dashed line in Fig. 3a. The top axis denotes  $n_e/(n_s/4)$  for  $x=0$  and the separation between the yellow dashed lines describes the evolution of  $n_s(x)$ . The

dispersive-band regions are marked in yellow. The signal in the flat bands is amplified seven times and multiplied by  $-1$  for p doping such that incompressible strips are bright. **b–d**, The derived position-dependent  $n_s(x)$  (**b**),  $\theta(x)$  (**c**) and the charge disorder  $\delta n_d(x)$  (**d**).

an out-of-plane magnetic field,  $B_a \approx 1$  T. We apply a small a.c. excitation,  $V_{\text{bg}}^{\text{ac}}$  onto the d.c. backgate voltage,  $V_{\text{bg}}^{\text{dc}}$ , which causes a small a.c. displacement  $\mathbf{r}_{\text{ac}}$  of the position of the  $I^{\text{T}}$  strips along the direction of the twist-angle gradient,  $\nabla\theta(\mathbf{r})$  (see Fig. 1a. Henceforth,  $\nabla = (\partial_x, \partial_y)$  represents the two-dimensional gradient in the  $x$ - $y$  plane). The corresponding a.c. Biot-Savart magnetic field,  $B_z^{\text{ac}}$ , is directly proportional to the local current density, eliminating the need for current reconstruction (see Methods).  $B_z^{\text{ac}}$  shows a sharp peak whenever the narrow (approximately 50-nm width) incompressible strips pass under the tip (Fig. 1a and Extended Data Fig. 4), providing very sensitive means for nanoscale imaging of the Landau levels.

Figure 1d shows a sequence of these  $B_z^{\text{ac}}$  peaks versus the local carrier density  $n_e$  for device B, acquired at a fixed SOT position, in comparison with the corresponding trace of the longitudinal resistance  $R_{xx}$  (Fig. 1c) at  $B_a = 1.08$  T. The position and magnitude of these peaks provide a wealth of information. An incompressible quantum Hall strip appears at location  $\mathbf{r}$  in the sample when the local carrier density precisely matches an integer number  $N$  of full Landau levels,  $|n_e(\mathbf{r})| = gN|B_a|/\phi_0$ , where  $g$  is the Landau level degeneracy and  $\phi_0 = h/e$  ( $e$  is the elementary charge,  $h$  is Planck's constant). Hence, the spacing  $\Delta n_e$  between adjacent peaks reveals the degeneracy  $g$  of the Landau levels. The height of the  $B_z^{\text{ac}}$  peaks is proportional to  $I^{\text{T}} = \sigma_{yx} \Delta \varepsilon_n / e$  (where  $\sigma_{yx} = ve^2/h$  is the quantum Hall conductance and  $v = gN$  is the integer filling factor), and thus reflects the energy gap between the adjacent Landau levels,  $\Delta \varepsilon_n = \varepsilon_{|n|+1} - \varepsilon_{|n|}$ , where  $n = 0, \pm 1, \pm 2, \dots$  is the Landau level index (see Methods).

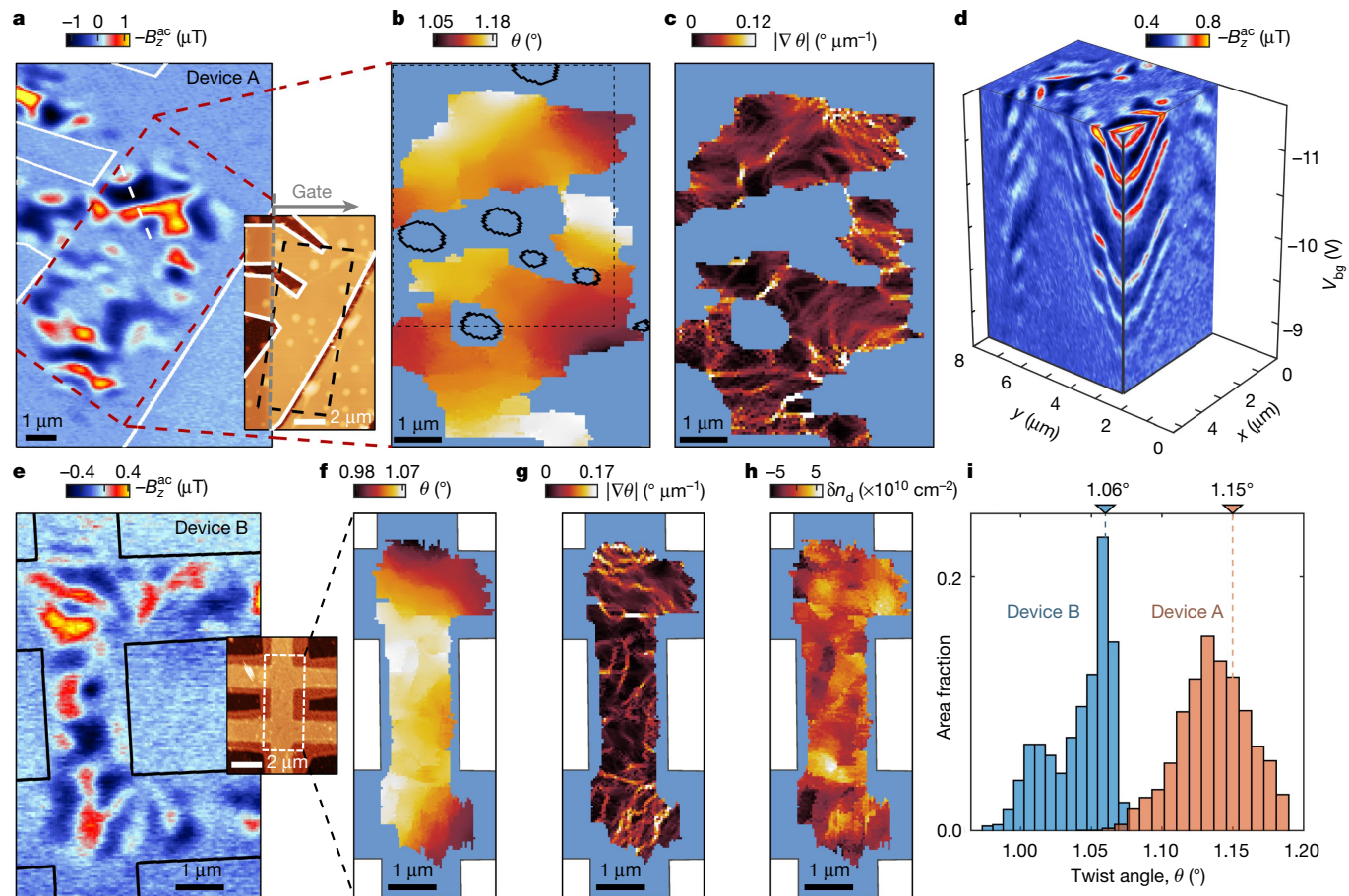
We start by inspecting high dopings,  $|n_e| > n_s$ , for which the Fermi level  $\varepsilon_F$  resides in the dispersive bands (yellow in Fig. 1e). Figure 1f presents a magnification of the four lowest Landau levels in the electron-like (n) and hole-like (p) dispersive bands for device A at  $B_a = 1.19$  T. The spacing between neighbouring peaks is  $\Delta n_e = 1.15 \times 10^{11} \text{ cm}^{-2} = 4B_a/\phi_0$ , showing that these Landau levels are fourfold degenerate. The spacing between the corresponding p and n Landau levels is  $2(n_e(\mathbf{r}) + 4N|B_a|/\phi_0)$ , as illustrated in Fig. 1f. The  $I^{\text{T}}$  peaks are very sharp, and so high-accuracy determination of the local  $n_s(\mathbf{r})$  and thus of the local twist angle

$\theta(\mathbf{r}) = a\sqrt{3n_s(\mathbf{r})}/8$  ( $a = 0.246$  nm is the graphene lattice constant) is obtained with an absolute accuracy of  $\pm 0.005^\circ$  and a relative accuracy between different locations of  $\pm 0.0002^\circ$  (see Methods). In the two-dimensional (2D) scanning mode described below, we attain a sensitivity of  $0.007^\circ$  per  $\text{Hz}^{1/2}$  and provide  $\theta(\mathbf{r})$  maps with relative accuracy better than  $\pm 0.002^\circ$ .

Instead of measuring at a fixed location, Fig. 2a shows  $B_z^{\text{ac}}$  in device A acquired upon scanning the SOT along the white dashed line in Fig. 3a and sweeping  $V_{\text{bg}}$ , revealing that the Landau levels vary in space, forming rich patterns. Moreover, the degeneracy of the higher Landau levels toggles between fourfold and eightfold as a function of position, and a pronounced asymmetry between the Landau level structure in the n and p dispersive bands is observed.

As in Fig. 1f, by tracing the spacing between the lower Landau levels we derive the local  $n_s(x) = C[V_{\text{ns}}(x) - V_{\text{ps}}(x)]/2$ , where  $V_{\text{ns}}(x)$  and  $V_{\text{ps}}(x)$  are the backgate voltages corresponding to the local filling of the flat bands  $|n_e(x)| = n_s(x)$  (dashed yellow curves in Fig. 2a), and  $C$  is the backgate capacitance (see Methods). The obtained  $n_s(x)$  (Fig. 2b) varies by about  $2.4 \times 10^{11} \text{ cm}^{-2}$ , corresponding to a local variation in  $\theta(x)$  of 3.9%, from  $1.124^\circ$  to  $1.169^\circ$  over the 2.7- $\mu\text{m}$ -long path (Fig. 2c). In addition to the twist-angle disorder, which shifts the p and n Landau levels antisymmetrically, we also derive the local charge disorder  $n_d(x)$ , which shifts all the Landau levels symmetrically through variation of the local charge neutrality point (CNP),  $n_d(x) = CV_{\text{CNP}}(x) = [V_{\text{ns}}(x) + V_{\text{ps}}(x)]/2$ . The derived charge disorder  $\delta n_d(x) = n_d(x) - \bar{n}_d$  (Fig. 2d) has a standard deviation (s.d.)  $= 0.8 \times 10^{10} \text{ cm}^{-2}$ , which is substantially smaller than the variation in  $n_s(x)$ , showing that the dominant source of disorder in this MATBG device arises from variations in  $\theta(\mathbf{r})$ , as evidenced in Fig. 2a by the antisymmetric bending of the dispersive p and n Landau levels.

To derive full maps of the local twist angle  $\theta(\mathbf{r})$  and the charge disorder  $\delta n_d(\mathbf{r})$ , we acquired  $B_z^{\text{ac}}(\mathbf{r})$  (Supplementary Videos 1–4) by incrementing  $V_{\text{bg}}$  through the bottom of the dispersive bands. Figure 3a displays one frame from Supplementary Video 2 showing a large-area



**Fig. 3 | Mapping the twist angle and Landau levels in MATBG.** **a**,  $B_z^{\text{ac}}$  image of the dashed area indicated in the inset at  $V_{\text{bg}} = -16.4$  V. Incompressible (compressible) quantum Hall regions are visible as bright blue to yellow (dark blue). Inset, AFM image of the hBN-encapsulated MATBG device A with edges outlined in white; the light-brown area indicates the underlying PdAu backgate and the bright spots show bubbles. **b**, 2D map of the twist angle  $\theta(\mathbf{r})$  derived from tomography of Supplementary Video 1 in the dashed region shown in **a**. Grey-blue indicates regions that do not display MATBG physics owing to disorder (bubbles outlined in black) or to a twist angle that is substantially

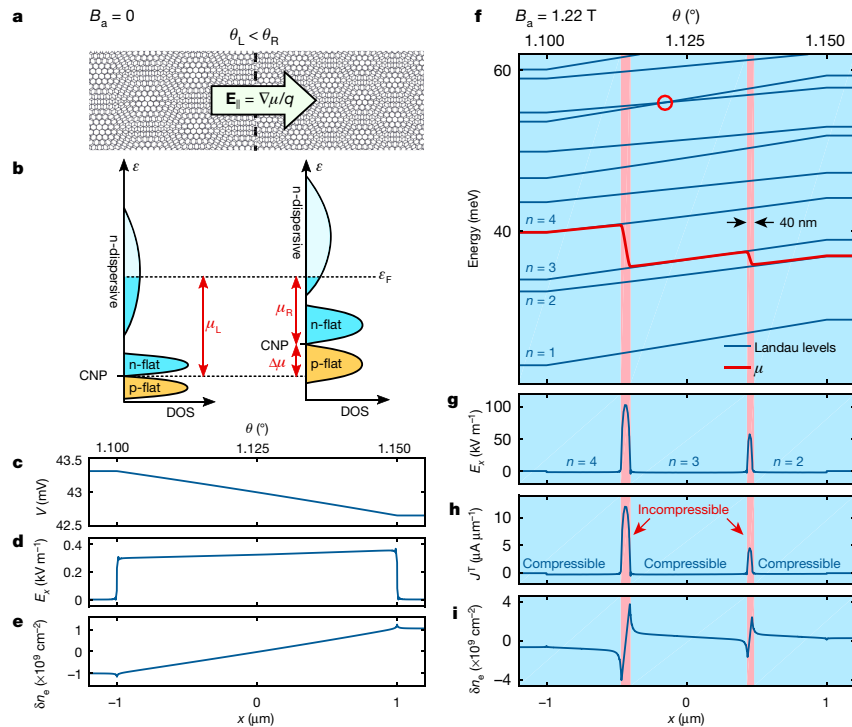
different to the magic angle. The dashed area is presented tomographically in **d**, **c**, 2D map of  $|\nabla\theta(\mathbf{r})|$  showing patches of slowly varying  $\theta(\mathbf{r})$  and a network of abrupt jumps in  $\theta$ . **d**, A slice from the tomography of device A showing disordered Landau levels in the bulk of the sample in the p dispersive band (Supplementary Video 5, and see ref. <sup>30</sup> for an interactive interface). The x-axis is flipped for clarity. **e–g**, As in **a–c**, for device B. **e** was acquired at  $V_{\text{bg}} = -15$  V and **f** is derived from tomography of Supplementary Videos 3, 4. **h**, Charge disorder map  $\delta n_d(\mathbf{r})$  of device B. **i**, Histogram of local  $\theta$  in devices A and B. The dashed lines mark the global  $\theta$  derived from transport measurements.

scan of device A (dashed rectangle in the inset), and Supplementary Video 1 presents a magnification of the central region (dashed rectangle in Fig. 3a). The red stripes reveal incompressible regions carrying  $I^{\text{T}}$  and the dark blue regions mark the compressible areas carrying the counterpropagating  $I^{\text{NT}}$ . As  $V_{\text{bg}}$  varies, the quantum Hall states move and change their shape in an intricate manner. Notably, the quantum Hall edge states are present in the bulk of the sample and do not flow parallel to the sample edges as would be expected. Moreover, large parts of the sample do not show Landau levels at all. These are the regions that are either highly disordered or may have a twist angle that is very different to the magic angle, with  $\theta$  either close to zero or  $\theta > 1.5^\circ$  such that the dispersive bands are reached at a  $V_{\text{bg}}$  outside our range. Thus, the magic-angle behaviour appears only in a limited central region of the sample and does not fully extend to the edges. By comparison, device B displays quantum Hall states over most of its area as revealed by the image of  $B_z^{\text{ac}}$  in Fig. 3e. Supplementary Videos 3 and 4 show the evolution of the Landau levels in the central part of the Hall bar structure (dashed rectangle in the inset) in the p and n dispersive bands, respectively.

Using these data we generate a three-dimensional (3D) tomographic rendering of the Landau levels throughout the samples (see Methods), which can be inspected interactively online<sup>30</sup>. Figure 3d shows a slice of the tomographic data of device A (see Supplementary Video 5),

revealing the layered structure of incompressible (light blue/red) and compressible (dark blue) quantum Hall regions. Notably, the Landau levels display steep slopes and numerous small jumps in the bulk of the sample, revealing that at any value of  $V_{\text{bg}}$  (represented by a horizontal tomographic plane) several different Landau levels cross  $\epsilon_{\text{f}}$  in the bulk of the sample, never forming a well defined single quantum Hall state. This observation explains the absence of clear conductance oscillations and quantization in the global  $R_{\text{xx}}$  data in the dispersive bands in Fig. 1c, despite the presence of fully developed Landau levels as observed locally in Fig. 1d.

Applying the procedure of Fig. 1f to the tomographic data, we derive 2D maps of the charge disorder (Fig. 3h; see discussion in Methods) and of the twist angle  $\theta(\mathbf{r})$  in devices A and B (Fig. 3b, f). The grey-blue colour in Fig. 3b reflects areas where no quantum Hall states were detected within the measured span of  $V_{\text{bg}}$ . These regions correlate with the locations of bubbles (black outlines) as revealed by atomic force microscopy (AFM) of device A (Fig. 3a, inset). Magic-angle behaviour is apparently absent within the bubbles and in their surrounding areas, up to  $0.5 \mu\text{m}$  from the bubble edges. The Landau levels are also absent in additional regions where no particular features were observed by AFM. The map in Fig. 3b also shows that the magic-angle regions in device A do not create a percolation path between the contacts. This is consistent with our transport measurements, which do not show



**Fig. 4 | Internal electric fields, unconventional quantum Hall state and equilibrium currents induced by twist-angle gradients.** **a**, Schematic representation of a MATBG sample with two connected regions of different twist angle, giving rise to an internal electric field  $E_{||}$ . **b**, Schematic DOS in the two regions with differing  $n_s$ . The backgate voltage  $V_{bg}$  imposes nearly the same electron density  $\bar{n}_e$  (blue areas) in the two regions; however, the difference in  $n_s$  results in different chemical potentials  $\mu_L$  and  $\mu_R$ , giving rise to  $E_{||} = \nabla\mu/q$ . **c–e**, Finite element calculation of the potential  $V(x) = -\mu(x)/q$  (**c**), the electric

field  $E_x(x)$  (**d**) and the carrier concentration  $\delta n_e = n_e(x) - \bar{n}_e$  (**e**) for the case of linear change in  $\theta$  from 1.10° to 1.15° with a 0.025°  $\mu\text{m}^{-1}$  gradient and an average carrier density  $\bar{n}_e = 3.25 \times 10^{12} \text{ cm}^{-2}$  at  $B_a = 0$ . **f–i**, Single-particle calculation of Landau level energies  $\varepsilon_n(x)$  in the electron-like dispersive band (blue) and the self-consistent calculation of the chemical potential  $\mu(x)$  (red; **f**),  $E_x(x)$  (**g**),  $J^||$  (**h**) and  $\delta n_e(x)$  (**i**) at  $B_a = 1.22 \text{ T}$ . Sharp peaks in  $E_x$  and  $J^||$  correspond to narrow (about 40-nm wide) incompressible strips (shaded red) that are observed experimentally as peaks in  $B_z^{\text{ac}}$ .

fully developed superconductivity, although correlated insulating states are present in this device. By contrast, in device B four-probe transport measurements showed high-quality correlated insulator states at multiple integer filling factors, and a zero-resistance superconducting state (see Methods) consistent with the observation that the magic-angle area extends over the entire length of the central part of the device and shows a more uniform  $\theta(\mathbf{r})$  (Fig. 3f).

The magic-angle regions show substantial twist-angle disorder (Fig. 3i, histogram).  $\theta(\mathbf{r})$  spans a range of 0.13° (1.05° to 1.18° with s.d. = 0.025°) in device A (Fig. 3b) and 0.10° in device B (0.98° to 1.08°, s.d. = 0.022°, Fig. 3f). Moreover, the topography of  $\theta(\mathbf{r})$  is nontrivial, with numerous peaks and valleys as well as saddle points. Because the Landau levels in the dispersive band follow the bottom of the band,  $n_s = 8\theta^2/(\sqrt{3}a^2)$ , they appear first at the minimum of the  $\theta(\mathbf{r})$  landscape, which for device A occurs in the lower-right corner (dark brown in Fig. 3b). This behaviour is clearly visible in Supplementary Video 1, in which arc-like incompressible strips (bright) first appear at this corner and upon increasing  $|V_{bg}|$  ‘climb’ the amphitheatre-like  $\theta(\mathbf{r})$  landscape following the equi- $\theta(\mathbf{r})$  contours. Similar behaviour is observed in other regions, with interesting dynamics occurring at the saddle points, as described in Methods.

The  $\theta(\mathbf{r})$  derived in Fig. 3b, f is smooth with typical gradients of around 0.05°  $\mu\text{m}^{-1}$ . Figure 3c, g (which shows the gradient maps  $|\nabla\theta(\mathbf{r})|$ ), reveals that variations in  $\theta(\mathbf{r})$  partially occur through a network of small steps of variable sizes, reaching a maximum step size of about 0.01°. The derived patterns strongly resemble the stacking-fault networks in bilayer graphene observed by transmission electron microscopy<sup>15,25–27</sup>. These steps cause the stepwise jumps in the Landau levels that are visible in the tomographic view in Fig. 3d and Extended Data Fig. 8a, c. This finding implies that the smooth variations in  $\theta(\mathbf{r})$  are accompanied by

occasional small, abrupt changes across stacking faults that relax the tensile and shear stress.

The revealed twist-angle disorder and gradients  $\nabla\theta(\mathbf{r})$  may have substantial implications on the phase diagram and transport properties of MATBG. Connecting regions of different  $\theta$  (Fig. 4a) is akin to connecting materials with different work functions (Fig. 4b), resulting in band-bending and creation of internal electric fields (Fig. 4c–e). The backgate voltage,  $V_{bg}$ , imposes a nearly uniform carrier density,  $\bar{n}_e \approx CV_{bg}$ . Locations with different  $\theta(\mathbf{r})$ , and hence a different density of states (DOS), translate this  $\bar{n}_e$  into a different chemical potential  $\mu(\mathbf{r})$ . At thermal equilibrium the Fermi level must be uniform,  $\varepsilon_F = \mu + qV = 0$  (where the last equality reflects the grounding of the device;  $q = \pm e$  is the charge of the carriers), and so variation in the chemical potential  $\mu(\mathbf{r})$  imposes band-bending—that is, it imposes variation in the electric potential  $V(\mathbf{r}) = -\mu(\mathbf{r})/q$  and creates an in-plane electric field  $E_{||} = -\nabla V$  that cannot be screened. Using the DOS derived from bandstructure calculations (see Methods), Fig. 4c–e presents a self-consistent numerical calculation of  $V(x)$ ,  $E(x)$  and  $\delta n_e(x) = n_e(x) - \bar{n}_e$  at  $B_a = 0$  for the case of linearly varying  $\theta(x)$  with  $\nabla\theta = 0.025^\circ \mu\text{m}^{-1}$  comparable to the measured average gradients in Fig. 3c, g. A large electric field  $E \approx 0.4 \text{ kV m}^{-1}$  is formed in the region of varying  $\theta(x)$ , whereas the accompanying charge redistribution remains negligible,  $\delta n_e(x)/\bar{n}_e \approx \pm 3 \times 10^{-5}$ . Note that for  $V_{bg} = 0$ , the twist-angle disorder has essentially no effect ( $\Delta\mu(\mathbf{r}) = 0$  and  $E_{||} = 0$ ) and its impact grows with increasing  $|\bar{n}_e|$ .

In the presence of a magnetic field, Landau levels are formed as depicted by the blue lines in Fig. 4f (see Methods). At high doping,  $|n_e| > n_s$ , gradients in  $\theta(\mathbf{r})$ —and therefore in  $n_s(\mathbf{r})$ —induce variation in the dispersive band occupation,  $|n_e| - n_s(\mathbf{r})$ , because  $n_e = CV_{bg}$  is approximately constant. As a result, an unusual quantum Hall state emerges



in which instead of being restricted to the edges, the quantum Hall edge states are formed in the bulk, creating interlaced compressible and incompressible strips with different integer filling factors (Fig. 4g–i). This absence of a well defined quantum Hall state provides an explanation for the Shubnikov–de Haas (SdH) oscillations without full conductance quantization that are commonly seen in MATBG magnetotransport<sup>1–7</sup>.

In contrast with conventional quantum Hall behaviour, in which the edge states must form closed loops, here the edge states seem to terminate in the bulk upon reaching apparently disordered metallic regions (Supplementary Videos 1–4). Moreover, instead of the constant carrier density that is usually required in incompressible regions, in the presence of a  $\theta$  gradient the density varies following the variation in  $n_s(\mathbf{r})$  (Fig. 4i).  $\nabla\theta(\mathbf{r})$  also causes accidental Landau level crossings (for example, indicated by the red circle in Fig. 4f) giving rise to occasional eightfold degenerate Landau levels in the dispersive bands, as observed in Figs. 1d, 2a and Extended Data Fig. 6. Figure 4g also shows large electric fields (approximately  $10^5 \text{ V m}^{-1}$ ) formed in the incompressible strips, giving rise to very narrow channels of persistent current  $I^T$  (Fig. 4h), consistent with the experimental data (Extended Data Fig. 4g). The typical width of the channels (around 50 nm) along with the local  $|\nabla\theta(\mathbf{r})|$ , determines the spatial resolution of our  $\theta(\mathbf{r})$  mapping (see Methods).

Finally, we discuss the rich structure observed in the flat bands in Figs. 1, 2. In contrast to transport measurements that resolve SdH oscillations at high fields where some of the degeneracies may be lifted, we probe the Landau levels locally at relatively low fields. The zeroth Landau level at the charge neutrality point is apparently eightfold degenerate, followed by fourfold degenerate Landau levels on both sides (Fig. 1d). It has been argued that such degeneracy indicates breaking of  $C_3$  symmetry<sup>31,32</sup>, which may in turn be triggered by the observed  $\theta$  gradients. Figures 1d, 2a show that these Landau levels are sometimes observed to extend beyond  $n_s/4$  on both the p and n sides, although at other locations new, irregular Landau levels seem to emerge for n doping above  $n_s/4$ , as visible in Fig. 1d. The Landau levels clearly reappear above  $n_s/2$  for both dopings, showing a degeneracy of 2 (Fig. 2a and Extended Data Fig. 7). We occasionally observe single-fold Landau levels above  $3n_s/4$  for both dopings, as seen in Fig. 1d and Extended Data Fig. 7. We also observe that the amplitudes of the  $I^T$  peaks stemming from different integer fillings often follow a smooth envelope. Since the amplitudes of the  $I^T$  peaks are proportional to the energy gaps  $\Delta\epsilon_n$ , this signifies that the energy gaps between consecutive Landau levels are of similar (not alternating) magnitudes, indicating full lifting of a degeneracy. Importantly, Figs. 1d, 2a and Extended Data Fig. 7 show that the Landau levels near  $n_s/4$  and  $n_s/2$  are discontinuous and that those above  $3n_s/4$  appear only at a few locations, indicating the extreme fragility of the correlated states to twist-angle disorder.

Twist-angle disorder as we have described here is a previously unrecognized type of disorder. It is fundamentally different to other, more common disorders, owing to its ability to change the local bandstructure and to induce large unscreened electric fields. The growth of its effect with the carrier density  $|n_e|$  explains the higher visibility of the Landau fan near the charge neutrality point in transport measurements. The charge disorder in graphene is commonly characterized by the width of the resistance peak at the charge neutrality point. This width, however, bears almost no information on twist-angle disorder. Instead, our results suggest that the twist-angle disorder should be quantified by analysing the width of the resistive peak at  $n_s$  and the visibility of the Landau fans at  $|n_e| > n_s$ . Our finding that the quantum Hall state is strongly altered by the twist-angle gradient suggests that other correlated phases in MATBG—including magnetism and superconductivity—may also be fundamentally transformed by the twist-angle disorder. The gate-tunable intrinsic in-plane electric fields generated by the twist-angle gradients may also be of practical importance for

photovoltaic and thermoelectric applications of atomically thin twisted van der Waals materials.

## Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-020-2255-3>.

1. Cao, Y. et al. Correlated insulator behaviour at half-filling in magic-angle graphene superlattices. *Nature* **556**, 80–84 (2018).
2. Cao, Y. et al. Unconventional superconductivity in magic-angle graphene superlattices. *Nature* **556**, 43–50 (2018).
3. Yankowitz, M. et al. Tuning superconductivity in twisted bilayer graphene. *Science* **363**, 1059–1064 (2019).
4. Serlin, M. et al. Intrinsic quantized anomalous Hall effect in a moiré heterostructure. *Science* **367**, 900–903 (2019).
5. Sharpe, A. L. et al. Emergent ferromagnetism near three-quarters filling in twisted bilayer graphene. *Science* **365**, 605–608 (2019).
6. Tomarken, S. L. et al. Electronic compressibility of magic-angle graphene superlattices. *Phys. Rev. Lett.* **123**, 046601 (2019).
7. Lu, X. et al. Superconductors, orbital magnets and correlated states in magic-angle bilayer graphene. *Nature* **574**, 653–657 (2019).
8. Vasyukov, D. et al. A scanning superconducting quantum interference device with single electron spin sensitivity. *Nat. Nanotechnol.* **8**, 639–644 (2013).
9. Uri, A. et al. Nanoscale imaging of equilibrium quantum Hall edge currents and of the magnetic monopole response in graphene. *Nat. Phys.* **16**, 164–170 (2020).
10. Suárez Morell, E., Correa, J. D., Vargas, P., Pacheco, M. & Barticevic, Z. Flat bands in slightly twisted bilayer graphene: tight-binding calculations. *Phys. Rev. B* **82**, 121407 (2010).
11. Bistritzer, R. & MacDonald, A. H. Moiré bands in twisted double-layer graphene. *Proc. Natl Acad. Sci. USA* **108**, 12233–12237 (2011).
12. Lopes dos Santos, J. M. B., Peres, N. M. R. & Castro Neto, A. H. Continuum model of the twisted graphene bilayer. *Phys. Rev. B* **86**, 155449 (2012).
13. Moon, P. & Koshino, M. Optical absorption in twisted bilayer graphene. *Phys. Rev. B* **87**, 205404 (2013).
14. Nam, N. N. T. & Koshino, M. Lattice relaxation and energy band modulation in twisted bilayer graphene. *Phys. Rev. B* **96**, 075311 (2017).
15. Yoo, H. et al. Atomic and electronic reconstruction at the van der Waals interface in twisted bilayer graphene. *Nat. Mater.* **18**, 448–453 (2019).
16. Huder, L. et al. Electronic spectrum of twisted graphene layers under heterostrain. *Phys. Rev. Lett.* **120**, 156405 (2018).
17. Bi, Z., Yuan, N. F. Q. & Fu, L. Designing flat bands by strain. *Phys. Rev. B* **100**, 035448 (2019).
18. Li, G. et al. Observation of van Hove singularities in twisted graphene layers. *Nat. Phys.* **6**, 109–113 (2010).
19. Brihuega, I. et al. Unraveling the intrinsic and robust nature of van Hove singularities in twisted bilayer graphene by scanning tunneling microscopy and theoretical analysis. *Phys. Rev. Lett.* **109**, 196802 (2012).
20. Wong, D. et al. Local spectroscopy of moiré-induced electronic structure in gate-tunable twisted bilayer graphene. *Phys. Rev. B* **92**, 155409 (2015).
21. Jiang, Y. et al. Flat bands in buckled graphene superlattices. Preprint at <https://arxiv.org/abs/1904.10147> (2019).
22. Kerelsky, A. et al. Maximized electron interactions at the magic angle in twisted bilayer graphene. *Nature* **572**, 95–100 (2019).
23. Choi, Y. et al. Electronic correlations in twisted bilayer graphene near the magic angle. *Nat. Phys.* **15**, 1174–1180 (2019).
24. Xie, Y. et al. Spectroscopic signatures of many-body correlations in magic-angle twisted bilayer graphene. *Nature* **572**, 101–105 (2019).
25. Alden, J. S. et al. Strain solitons and topological defects in bilayer graphene. *Proc. Natl Acad. Sci. USA* **110**, 11256–11260 (2013).
26. Lin, J. et al. AC/AB stacking boundaries in bilayer graphene. *Nano Lett.* **13**, 3262–3268 (2013).
27. Butz, B. et al. Dislocations in bilayer graphene. *Nature* **505**, 533–537 (2014).
28. Cao, Y. et al. Superlattice-induced insulating states and valley-protected orbits in twisted bilayer graphene. *Phys. Rev. Lett.* **117**, 116804 (2016).
29. Kim, K. et al. van der Waals heterostructures with high accuracy rotational alignment. *Nano Lett.* **16**, 1989–1995 (2016).
30. Landau Level Tomography of Magic Angle Twisted Bilayer Graphene (MATBG) (2019); [www.weizmann.ac.il/condmat/superc/software/matbg](http://www.weizmann.ac.il/condmat/superc/software/matbg).
31. Hejazi, K., Liu, C. & Balents, L. Landau levels in twisted bilayer graphene and semiclassical orbits. *Phys. Rev. B* **100**, 035115 (2019).
32. Zhang, Y. H., Po, H. C. & Senthil, T. Landau level degeneracy in twisted bilayer graphene: role of symmetry breaking. *Phys. Rev. B* **100**, 125104 (2019).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2020

## Methods

### Device fabrication

The MATBG devices were fabricated using the previously reported ‘tear and stack’ technique<sup>28,29,33</sup>. We first exfoliated monolayer graphene and hBN of 10- to 50-nm thickness on SiO<sub>2</sub>/Si substrates, annealed at 350 °C (for hBN only) and selected using optical microscopy and AFM. Only flakes without wrinkles and bubbles were used. A PC/PDMS polymer stack on a glass slide mounted on a micro-positioning stage was used to pick up an approximately 10-nm-thick hBN flake. The edge of the hBN flake was then used to tear a graphene flake. The substrate was rotated by 1.1° to 1.2°, followed by pickup of the other piece of graphene. The resulting stack was encapsulated with another hBN flake of thicknesses of 30–70 nm that had been put onto a metallic gate made of evaporated Cr/PdAu. The device geometry was defined by electron-beam lithography and reactive ion etching, keeping only the relatively clean regions. Electrical contacts to the MATBG were made by the one-dimensional edge-contact method<sup>34</sup>.

Optical images of devices A and B are shown in Extended Data Fig. 1a, b, respectively. Device A was fabricated on a degenerately doped Si substrate with 300-nm SiO<sub>2</sub>. The MATBG resides partly on SiO<sub>2</sub> and partly on the evaporated metallic backgate (light brown in Extended Data Fig. 1a). In this work, only the metallic backgate has been used for varying the carrier concentration  $n_e$  and a constant voltage  $V_{bg}^{Si} = 50$  V was applied to the Si backgate for keeping the rest of the sample conductive during the transport measurements. Device B was fabricated on an intrinsic Si substrate with a metallic backgate extending over the full size of the device (light blue in Extended Data Fig. 1b).

### Transport characteristics

Four-probe resistance measurements of the samples at  $T = 300$  mK are shown in Extended Data Figs. 2, 3. Both devices exhibit the common transport characteristics of correlated physics in MATBG<sup>1–3,5–7</sup>, including  $R_{xx}$  peaks at  $n_e$  and its integer fractions, and Landau fans at elevated magnetic field. The slopes of the Landau fans in Extended Data Fig. 2a, b were used to extract the backgate capacitances  $C$  of  $3.07 \times 10^{11} \text{ cm}^{-2} \text{ V}^{-1}$  ( $49.23 \text{ nF cm}^{-2}$ ) for device A and  $2.31 \times 10^{11} \text{ cm}^{-2} \text{ V}^{-1}$  ( $37 \text{ nF cm}^{-2}$ ) for device B, consistent with the evaluated dielectric thickness of the underlying hBN. The origins of the Landau fans were used to derive the global  $n_s$  and the corresponding global  $\theta = 1.15^\circ$  for device A and  $\theta = 1.06^\circ$  for device B, in good correspondence with histograms of the local twist angle in Fig. 3i. In device A the global  $\theta$  correlates with the average of the  $\theta(\mathbf{r})$  distribution, whereas in device B it is close to the upper end of the distribution function. Consistently, the four-probe transport measurements in device B probe the central part of the Hall bar structure (Extended Data Fig. 1b), where  $\theta(\mathbf{r})$  is at its highest and is considerably more uniform (Fig. 3f), whereas the low end tail of  $\theta(\mathbf{r})$  distribution arises from regions that are not probed by transport measurements.

In addition, in device B we observe the superconducting state in the vicinity of p-doped  $n_s/2$  with zero  $R_{xx}$ ; this superconducting state becomes suppressed by small magnetic field (Extended Data Fig. 3b). The critical current in the superconducting state reaches about 100 nA, as determined by the differential  $dI/dI$  characteristics (Extended Data Fig. 3c), and depends sensitively on the carrier density  $n_e$ . The observation of fully developed superconductivity in device B is consistent with the finding of a continuous magic-angle region between the voltage contacts in Fig. 3f. Suppression of the resistance was also observed in device A (Extended Data Fig. 3a), but the lowest  $R_{xx}$  was 328  $\Omega$ , suggestive of the presence of some superconducting regions but absence of a percolation path between the voltage contacts, consistent with the  $\theta$  map in Fig. 3b.

### SOT fabrication and characterization

The Pb SOTs were fabricated as described in ref. <sup>8</sup> with diameters ranging from 220 to 250 nm, and they included an integrated shunt resistor

on the tip<sup>35</sup>. The SOT readout was carried out using a cryogenic SQUID series array amplifier<sup>36–38</sup>. The magnetic imaging was performed in a <sup>3</sup>He system<sup>39</sup> at 300 mK, at which the Pb SOTs can operate in magnetic fields of up to 1.8 T. At the fields  $B_a \approx 1.2$  T used in this study, the SOTs displayed flux noise down to 250 n $\Phi_0$  per Hz<sup>1/2</sup> (where  $\Phi_0$  is the superconducting flux quantum,  $\Phi_0 = h/2e \approx 2 \times 10^{-7} \text{ G cm}^2$ ), spin noise of 10  $\mu_B$  per Hz<sup>1/2</sup> ( $\mu_B$ , Bohr magneton), and field noise down to 10 nT per Hz<sup>1/2</sup>. For height control, we attached the SOT to a quartz tuning fork, as described in ref. <sup>40</sup>. The tuning fork was electrically excited at the resonance frequency of ~33 kHz. The current through it was amplified using a room-temperature custom-built trans-impedance amplifier, designed based on ref. <sup>41</sup> and measured using a lock-in amplifier. The scanning was performed at a constant height of 20 to 100 nm above the surface of the top hBN.

### Technique for direct imaging of the current and evaluation of the current in the incompressible strips

To avoid the  $1/f$  noise of the SOT that is present at frequencies below ~1 kHz, an a.c. signal from backgate modulation was acquired instead of measuring the local d.c.  $B_z(\mathbf{r})$ . We applied a small a.c. excitation to the backgate (Fig. 1a),  $V_{bg} = V_{bg}^{dc} + V_{bg}^{ac} \sin(2\pi ft)$ , where  $f \approx 3$  kHz, and the corresponding  $B_z^{ac} = V_{bg}^{ac} \partial B_z / \partial V_{bg}$  was then measured by the SOT using a lock-in amplifier. Another major advantage of this modulation is that it provides a convenient method for direct imaging of the local current density  $J(\mathbf{r})$ . To demonstrate its principle, consider a  $\theta$  gradient in the  $x$  direction that gives rise to a narrow strip of current of width  $\Delta x$  positioned at  $x_0$ , and carrying a current density  $J_y$  in the  $y$  direction with a total current  $I_y = \Delta x J_y$  (Extended Data Fig. 4a). The magnetic field  $B_z(x)$  generated by the current and measured at a height  $h_{SOT}$  above it is described by the Biot–Savart law (Extended Data Fig. 4b). For heights  $h_{SOT} > \Delta x$ ,  $B_z(x)$  is largely governed by only the total current  $I_y$  in the strip, independent of  $\Delta x$ .  $B_z(x)$  is an antisymmetric function with a steep slope above the current strip. Its spatial derivative  $\partial B_z / \partial x$  has a sharp peak at the strip location (Extended Data Fig. 4c), with a height proportional to  $I_y$ , and thus can provide a good means for direct imaging of the current-density distribution  $J_y(x)$  if this distribution can be modulated in space in the  $x$  direction. The backgate voltage  $V_{bg}^{ac}$  provides this spatial modulation, as follows. In the presence of potential gradients, the quantum Hall edge channels flow along equipotential contours (given by equi- $\theta$  contours in the absence of charge disorder). A small  $V_{bg}^{ac}$  thus shifts the location of the channel by  $x_0^{ac} = V_{bg}^{ac} \partial x_0 / \partial V_{bg}$  in the direction parallel to the gradient and perpendicular to the current flow. So regardless of the gradient direction  $\hat{x}$ , the measured signal will be given by  $B_z^{ac} = -x_0^{ac} \partial B_z / \partial x \propto x_0^{ac} J_y(x)$ , thus providing direct imaging of the local current density. Extended Data Fig. 4d–f presents a simulation of three counterpropagating current strips demonstrating the  $B_z^{ac}$  imaging for this case.

The sharpness of the  $B_z^{ac}$  peak is determined by  $\Delta x$ ,  $h_{SOT}$ ,  $x_0^{ac}$  and the SOT diameter. In Fig. 2a, the scanning height above the MATBG is  $h_{SOT} = 70$  nm (including hBN) and the effective SOT diameter is 220 nm; hence the spatial resolution is largely determined by the SOT diameter. For these parameters and a root mean square (r.m.s.)  $x_0^{ac} = 54$  nm, Extended Data Fig. 4c shows that a current strip that is narrower than ~150 nm will result in a resolution-limited  $B_z^{ac}$  peak (compare the solid and dashed lines).

We now use this numerical procedure to analyse the data in Fig. 2a, which presents the  $B_z^{ac}$  signal along the white dashed line in Fig. 3a; this dashed line is aligned along the twist-angle gradient in a region of largely smooth  $\theta(\mathbf{r})$  behaviour (light brown region in the top-left corner of Fig. 3b, see also Supplementary Video 1). Extended Data Fig. 4g presents an example of the  $B_z^{ac}$  data from Fig. 2a at  $V_{bg} = -10.54$  V displaying the  $B_z^{ac}$  peak (blue). The data at consecutive  $V_{bg}$  values show that the peak position  $x_0$  shifts with  $V_{bg}$  at a rate of  $\partial x_0 / \partial V_{bg} = 1.54 \mu\text{m V}^{-1}$ , corresponding to a twist-angle gradient  $\frac{\partial \theta}{\partial x} = C / \left( \frac{\partial x_0}{\partial V_{bg}} \frac{\partial n_s}{\partial \theta} \right) = 0.0374^\circ \mu\text{m}^{-1}$

# Article

(where  $\frac{\partial n_s}{\partial \theta} = \frac{16}{\sqrt{3}} \frac{\theta}{a^2}$ ). The data were acquired with an r.m.s. voltage,  $V_{bg}^{ac} = 35$  mV, inducing an r.m.s.  $x_0^{ac} = V_{bg}^{ac} \frac{\partial x_0}{\partial V_{bg}} = 54$  nm. The red curve in Extended Data Fig. 4g shows a numerical fit to the data using these experimental parameters and a current-strip width of  $\Delta x = 50$  nm, resulting in  $I^T = 1.3$   $\mu$ A. The good fit shows that the experimental results are consistent with our simulations presented in Fig. 4f–i, although the exact value of  $\Delta x$  cannot be determined because the experimental  $B_z^{ac}$  peak is resolution-limited by the diameter of the SOT.

The extracted value  $I^T = 1.3$   $\mu$ A is not sensitive to precise  $\Delta x$ . Because  $I^T = \sigma_{yx} \Delta \varepsilon_n / e = v \Delta \varepsilon_n e / h$ , and  $v = -12$  in Extended Data Fig. 4g, we obtain  $\Delta \varepsilon_n = 2.8$  meV, comparable to the values derived from bandstructure calculations (Extended Data Fig. 10a). Using  $\Delta x = 50$  nm we obtain (in the incompressible region) an in-plane electric field of  $E_{||} = \Delta \varepsilon_n / (e \Delta x) = 56$  kV m<sup>-1</sup> comparable to the simulation values in Fig. 4g.

## Topological and nontopological currents in the incompressible and compressible quantum Hall strips

Gradients in the twist angle  $\nabla \theta$  give rise to gradients in the chemical potential  $\nabla \mu$  and to alternating compressible (when  $\mu$  resides within a Landau level) and incompressible ( $\mu$  in the energy gap between Landau levels) quantum Hall strips (Fig. 4f–i). Both regions carry current<sup>42</sup>; however, usually only the currents in the incompressible strips,  $\mathbf{J}^I = \sigma \mathbf{E}$ , which are of topological nature, are considered, whereas the nontopological currents in the compressible strips,  $\mathbf{J}^{NT} = \mu_e \nabla \times |\mathbf{n}_e| \hat{\mathbf{z}}$ , are commonly ignored (here  $\mu_e = e_k / B$  is the magnetic moment of the orbiting electron and  $e_k$  is its kinetic energy)<sup>43</sup>. The following semiclassical picture is instructive in describing  $\mathbf{J}^I$  and  $\mathbf{J}^{NT}$ . Under strong magnetic fields and in the absence of in-plane electric fields, the charge carriers follow cyclotron orbits that can be described semiclassically as an array of circles, resulting in zero average bulk current (Extended Data Fig. 5a). Applying an external in-plane electric field along the  $x$  direction to an incompressible state ( $E_x = -\partial V / \partial x$ ) causes the circular orbitals to convert into spirals drifting along the  $y$  direction, generating a current  $J_y^I = \sigma_{yx} E_x$  (Extended Data Fig. 5b). Alternatively, applying the same external electric field to a compressible strip will result in carrier redistribution that screens the in-plane electric field. As a result the drift current vanishes, but at the cost of a non-zero gradient in the carrier density,  $\partial n_e / \partial x$  (Extended Data Fig. 5c). Because each orbital carries a magnetic moment  $\mu_e = \mu_e \hat{\mathbf{z}}$ , giving rise to a local magnetization  $\mathbf{m} = |\mathbf{n}_e| \mu_e$ , the induced  $\partial n_e / \partial x$  causes gradients in  $\mathbf{m}$ , and hence produces equilibrium currents through<sup>42</sup>  $\mathbf{J}^{NT} = \nabla \times \mathbf{m}$ . This accounts for a non-zero  $J_y^{NT} = \mu_e \partial |\mathbf{n}_e| / \partial x$  (cyan arrows in Extended Data Fig. 5c), which flows in the direction opposite to the topological current  $J_y^I$  in Extended Data Fig. 5b. Since a full band does not contribute to current,  $n_e$  in the above expression refers only to carriers in a partially filled band. Alternatively,  $J_y^{NT}$  can be understood as arising from uncompensated contributions to the current from neighbouring orbitals in the presence of a gradient in the orbital density (Extended Data Fig. 5c).

The total current carried by the drifting orbitals in an incompressible strip residing between two compressible regions is given by  $I_y^I = \int J_y^I dx = \sigma_{yx} \Delta \varepsilon_n / e$ , where  $\Delta \varepsilon_n = \varepsilon_{|n|+1} - \varepsilon_{|n|}$  is the Landau level energy gap between adjacent compressible states and  $\sigma_{yx} = ve^2 / h$  is the quantum Hall conductance of the incompressible state (see Fig. 4). For a more extensive description of  $\mathbf{J}^I$  and  $\mathbf{J}^{NT}$  see ref.<sup>9</sup>.

## Determination of twist-angle measurement accuracy and spatial resolution

**$\theta$  accuracy.** The local twist angle is determined by the local  $n_s(\mathbf{r})$  via  $\theta(\mathbf{r}) = a \sqrt{3} n_s(\mathbf{r}) / 8$ . The incompressible  $I^I$  current and the corresponding peak in the  $B_z^{ac}$  signal appear at specific locations where  $N$  Landau levels in the dispersive bands ( $N = 1, 2, 3, \dots$ ) are exactly fully occupied, corresponding to a density  $|n_e| = C |V_{bg}^N - V_{bg}^{CNP}| = n_s + 4N|B_a| / \phi_0$  for four-fold degenerate Landau levels, where  $V_{bg}^N$  is the backgate voltage that

corresponds to the  $N$ th peak. Measuring the  $N^-$  and  $N^+$  peaks in the  $p$  and  $n$  dispersive bands, respectively, allows derivation of  $n_s(\mathbf{r}) = C(|V_{bg}^{N+}| + |V_{bg}^{N-}|) / 2 - 4N|B_a| / \phi_0$  and therefore of  $\theta(\mathbf{r})$ . The absolute angle accuracy is thus determined by the accuracy of  $C$ ,  $B_a$  and  $V_{bg}^{N\pm}$ . Determining  $C$  is possible through global transport measurements and more accurately through local measurement of the spacing between any two incompressible peaks  $V_{bg}^{N+1} - V_{bg}^N = g|B_a| / (\phi_0 C)$ , where  $g$  is the degeneracy of the Landau level considered. From this we estimate our overall absolute accuracy of determining  $n_s$  to be about  $\pm 1\%$ , thus giving an absolute  $\theta$  accuracy of  $\delta \theta = \pm 0.005^\circ$ .

In this study, however, we are particularly interested in the relative accuracy of  $\theta(\mathbf{r})$  for comparing different locations  $\mathbf{r}$  and deriving the angle gradients  $\nabla \theta$ . The relative accuracy is determined primarily by the measurement precision of  $V_{bg}^{N\pm}$ . The sharpness of the  $I^I$  peaks and the good signal-to-noise ratio of the  $B_z^{ac}$  signal allow high-precision measurement of  $V_{bg}^{N\pm}$  as demonstrated in Extended Data Fig. 6. In the stationary measurement in Extended Data Fig. 6a (magnification of Fig. 1f),  $V_{bg}$  was swept with increments  $\Delta V_{bg} = 4.7$  mV, demonstrating that the  $V_{bg}^{N=-3}$  and  $V_{bg}^{N=-4}$  peak positions can be determined to an accuracy better than  $\pm \Delta V_{bg}$ , (one step size), corresponding to  $\delta V_{bg}^{N=-4} / V_{bg}^{N=-4} \approx 4 \times 10^{-4}$ .  $\theta \propto \sqrt{n_s}$ , and so we have  $\delta \theta / \theta \approx 2 \times 10^{-4}$ , or a relative  $\theta$  accuracy of  $\delta \theta = \pm 0.0002^\circ$ . In Supplementary Video 1 (which was used to construct the full  $\theta(\mathbf{r})$  map of device A; see Fig. 3b), larger increments  $\Delta V_{bg} = 40$  mV were used (Extended Data Fig. 6b), corresponding to a  $\theta$  accuracy of  $\delta \theta = \pm 0.002^\circ$ . Supplementary Videos 3, 4 used  $\Delta V_{bg} = 45$  mV constructing the  $\theta(x, y)$  map of device B (Fig. 3f) with similar accuracy. Supplementary Videos 3, 4 contain 87 frames of  $68 \times 184 = 12,512$  pixels each, which were acquired over a total of 42 h. The  $V_{bg}$  trace of each pixel therefore took  $t = 12$  s to acquire. The  $\delta \theta \leq \pm 0.002^\circ$  accuracy, normalized by the pixel acquisition time, indicates that the relative  $\theta$  sensitivity per pixel in the imaging mode is better than  $S_\theta^{1/2} = \sqrt{t} \delta \theta = 0.007^\circ$  per Hz<sup>1/2</sup>.

**Spatial resolution of  $\theta(\mathbf{r})$  mapping.** Our electrostatic simulations show that the typical width of the incompressible  $I^I$  strips is about 50 nm (Fig. 4f, h) and should be smoothed by the wavefunction width, of the order of magnetic length  $l_b = \sqrt{\hbar / eB} \approx 25$  nm. The position  $\mathbf{r}$  of the incompressible strip provides a very accurate determination on the local  $n_s(\mathbf{r})$  and  $\theta(\mathbf{r})$ , and so the width of the strip essentially determines the spatial resolution, which can be smaller than the SOT diameter. The actual spatial resolution  $\delta r$  is determined by the accuracy  $\delta V_{bg}$  to which the  $V_{bg}$  value can be assigned to the  $I^I$  peak at a location  $\mathbf{r}$ ,  $\delta r = \delta V_{bg} dr / \partial V_{bg}$ , where  $dr / \partial V_{bg}$  is the change in the position of  $I^I$  per change in  $V_{bg}$ . Since  $I^I$  appears at  $V_{bg}(\mathbf{r}) = [n_s(\mathbf{r}) + 4N|B_a| / \phi_0] / C$ , the space dependence enters only through  $n_s(\mathbf{r}) = n_s(\theta(\mathbf{r}))$ , and thus  $dr / \partial V_{bg} = C(\partial \theta / \partial r)^{-1} (\partial n_s / \partial \theta)^{-1}$ , where  $\partial n_s / \partial \theta = 16\theta / \sqrt{3} a^2$ . Using characteristic values  $C = 2.5 \times 10^{11}$  V<sup>-1</sup> cm<sup>-2</sup>,  $\delta V_{bg} = 45$  mV in the scanning mode and  $\partial \theta / \partial r = 0.05^\circ \mu\text{m}^{-1}$  gives a resolution  $\delta r = 50$  nm. Smaller  $\partial \theta / \partial r$  gradients result in larger  $\delta r$ . However, in such a case, a lower spatial resolution is required because  $\theta$  varies slowly in space. The estimated  $\delta r$  is comparable to the pixel size in the videos (57 nm in Supplementary Video 1 and 43 nm in Supplementary Videos 3, 4). We thus conclude that the spatial resolution  $\delta r$  of the attained  $\theta(\mathbf{r})$  maps is of the order of 4–5 moiré supercells (13 nm each).

## Local quantum Hall measurement in device A

Extended Data Fig. 7 presents the local  $B_z^{ac}$  measurement with the SOT at a fixed position, along with the global transport  $R_{xx}$  measurement in device A at  $B_a = 1.19$  T. Alternating compressible and incompressible states in the region under the tip leads to a series of peaks in  $B_z^{ac}$ , with sharp peaks corresponding to incompressible strips carrying  $I^I$ . The sign of the incompressible peaks is determined by the sign of  $\sigma_{yx}$ , with  $B_z^{ac} > 0$  ( $B_z^{ac} < 0$ ) for electron (hole) doping. In Figs. 1d, 2a the  $B_z^{ac}$  signal for  $p$  doping is multiplied by  $-1$  for clarity. The spacing between adjacent



peaks reflects the degeneracy of the Landau level. The dispersive band (shaded yellow), exhibits a sequence of fourfold and eightfold degeneracies. In the flat band we find fourfold degenerate levels around  $n_e = 0$ , twofold degeneracy near  $n_e = \pm n_s/2$ , and 1-fold degenerate levels near  $n_e = -3n_s/4$  (see Fig. 1d for 1-fold degenerate levels near  $n_e = +3n_s/4$ ). Evaluation of the local  $n_s$  allows the extraction of the local twist angle,  $\theta = 1.136 \pm 0.005^\circ$ , as described below. In contrast to the sharp local  $B_z^{\text{ac}}$  peaks, oscillations in  $R_{xx}$  are barely visible, owing to  $\theta(\mathbf{r})$  disorder and the fact that the magic-angle regions in device A do not extend over the entire device area.

### Landau level tomography and twist-angle mapping

To map the local twist angle, a series of  $B_z^{\text{ac}}(\mathbf{r})$  area scans were performed while varying  $V_{\text{bg}}$ . This results in a 3D dataset with two spatial dimensions and one  $V_{\text{bg}}$  (or equivalently,  $n_e$ ) axis. Each energy gap between adjacent Landau levels forms a 2D manifold in this 3D space with a peak in the  $B_z^{\text{ac}}$  signal (bright regions in Extended Data Fig. 8). The manifolds of the lowest Landau levels in the dispersive bands trace the manifold of the bottom of the dispersive band,  $n_s(\mathbf{r})$ , and are displaced vertically from it by the degeneracy of the Landau levels, thus providing the means for mapping the local  $n_s(\mathbf{r})$  and hence the local  $\theta(\mathbf{r}) = a\sqrt{3}n_s(\mathbf{r})/8$ . The 3D space was mapped with pixel size of  $\sim 50$  nm and  $V_{\text{bg}}$  spacing between successive scans,  $\Delta V_{\text{bg}} \approx 40$  mV, which enables us to map  $\theta(\mathbf{r})$  with an accuracy of  $\delta\theta = \pm 0.002^\circ$  (see Methods).

For device A, the tomographic imaging was acquired for the p dispersive band for  $V_{\text{bg}}$  spanning  $-8.58$  V to  $-11.50$  V with  $\Delta V_{\text{bg}} = 40$  mV (Supplementary Video 1). The spacing between adjacent fourfold levels at  $B_s = 1.22$  T was  $0.39$  V  $\approx 10\Delta V_{\text{bg}}$ . In this device, the spatial variation of the charge neutrality voltage  $V_{\text{bg}}^{\text{CNP}}(\mathbf{r})$  was found to be very small (Fig. 2a) and therefore  $n_s(\mathbf{r})$  was derived from the 3D data assuming a constant  $V_{\text{bg}}^{\text{CNP}}$ . Representative slices of the 3D dataset are shown in Extended Data Fig. 8a, b. At  $V_{\text{bg}} = -8.5$  V, the Fermi level resides in the flat band for all points in space, and at  $V_{\text{bg}} = -11.5$  V,  $\varepsilon_F$  is in the dispersive band. As  $\varepsilon_F$  moves through the bottom of the dispersive band, it crosses four fourfold degenerate Landau levels above  $n_s$ , followed by an eightfold degenerate Landau level. The black line in Extended Data Fig. 8a traces the  $N = -4$  incompressible  $I^T$  peak revealing gradients with occasional small jumps in the twist angle. We note that at the jump positions the intensity of the signal is suppressed, owing to pinning of the Landau levels at the steps, which reduces the amplitude of the spatial a.c. displacement  $x_0^{\text{ac}}$  and hence the intensity of  $B_z^{\text{ac}}$  (see Methods section ‘Technique for direct imaging of the current and evaluation of the current in the incompressible strips’).

We note that at any value of  $|V_{\text{bg}}| > n_s/C$  several different Landau levels cross  $\varepsilon_F$  in the bulk of the sample. Thus, despite the fact that fully developed Landau levels are present locally, no well defined quantum Hall state in the dispersive bands can be observed globally. The same is true for the Landau levels of the correlated states, which are also influenced by variations in  $n_s(\mathbf{r})$ . This explains why MATBG magnetotransport commonly shows SdH oscillations without displaying full conductance quantization, with the exception of the CNP Landau fan<sup>1-7</sup>. At high enough field, the quantum Hall quantization should be recovered when the Landau level degeneracy  $4B_s/\phi_0$  exceeds the  $n_s(\mathbf{r})$  variations.

Device B exhibited stronger charge inhomogeneity and hence the 3D tomographic imaging was acquired for both the p and n dispersive bands (Supplementary Videos 3, 4) and  $n_s(\mathbf{r})$  was derived from the separation between the corresponding Landau levels in the two bands as described schematically in Fig. 1f. The tomographic data of both samples has been published online<sup>30</sup>.

The observation of superconductivity in devices with twist-angle disorder of  $\sim 0.1^\circ$  may be explained by either a tolerance of the superconducting state to the exact  $\theta$  or by percolating paths along very specific  $\theta_M$ . Both of our devices show superconductivity—despite having only a small overlap in their histograms in Fig. 3i—which supports the former

explanation. Figure 2a shows, however, that Landau levels near  $n_s/4$  and  $n_s/2$  are discontinuous and those above  $3n_s/4$  appear only at a few locations, indicating the extreme fragility of the correlated states to twist-angle disorder.

### Mapping of the charge disorder

Similarly to the mapping of the twist angle disorder through  $n_s(\mathbf{r}) = C[V_{\text{ns}}(\mathbf{r}) - V_{-\text{ns}}(\mathbf{r})]/2$ , the tomographic imaging also allows mapping of the charge disorder  $\delta n_d(\mathbf{r}) = n_d(\mathbf{r}) - \bar{n}_d$ , where  $n_d(\mathbf{r}) = C[V_{\text{ns}}(\mathbf{r}) + V_{-\text{ns}}(\mathbf{r})]/2$ , as presented in Fig. 3h for device B. Extended Data Fig. 9 shows the histogram of  $\delta n_d(\mathbf{r})$  along with a Gaussian fit with a standard deviation  $\Delta n_d = 2.59 \times 10^{10} \text{ cm}^{-2}$ , which is comparable to high-quality hBN-encapsulated monolayer graphene devices<sup>44</sup> and substantially lower than in graphene on  $\text{SiO}_2$ <sup>45</sup>. Note that in contrast to hBN-encapsulated graphene, the MATBG fabrication process is currently incompatible with thermal annealing procedures for disorder reduction. We observe that the charge disorder in device B is notably larger than in the magic-angle regions in device A (Fig. 2d), which we ascribe to the fact that in contrast to device A, device B did not undergo surface residue cleaning by AFM.

Note that the tomographic method allows mapping of the twist-angle and charge disorders only in the magic-angle regions, where Landau levels are present. In device A, a large part of the sample did not show magic-angle physics (Supplementary Videos 1, 2 and Fig. 3a, b), whereas the magic-angle regions revealed very low charge disorder with an estimated standard deviation  $\Delta n_d \approx 1.3 \times 10^{10} \text{ cm}^{-2}$  as attained by several 1D scans, such as in Fig. 2a, d. We therefore performed tomographic imaging of only the p dispersive band, which does not permit extraction of the full 2D map of  $\delta n_d(\mathbf{r})$  in device A. Neglecting this low level of charge disorder introduces an error in the derived  $\theta(\mathbf{r})$  map of device A of  $\delta\theta \leq 0.0015^\circ$ , which is negligible compared to the span of  $\theta(\mathbf{r})$  in Fig. 3b.

### Bandstructure calculations and Landau level crossings

The bandstructure of twisted bilayer graphene can be computed from an effective continuum Hamiltonian<sup>11,12,46-48</sup>:

$$H^{(\xi)} = \begin{pmatrix} H_1^{(\xi)} & U^\dagger \\ U & H_2^{(\xi)} \end{pmatrix}$$

where  $H_i^{(\xi)}$  is the valley-dependent monolayer graphene Hamiltonian for layer  $i$ :

$$H_i^{(\xi)} = -\hbar v_F(k - K_i^{(\xi)}) \cdot (\xi \sigma_x, \sigma_y)$$

with Fermi velocity  $v_F$ ,  $\xi = \pm 1$  indicating the positive and negative valleys,  $K_i^{(\xi)}$  the  $k$ -space location of the respective Dirac points in layer  $i$ , and  $U$  the interlayer coupling<sup>48-51</sup>:

$$U = \begin{pmatrix} u & u' \\ u' & u \end{pmatrix} + \begin{pmatrix} u & u'\omega^* \\ u'\omega & u \end{pmatrix} e^{i\xi \mathbf{G}_1^m \cdot \mathbf{r}} + \begin{pmatrix} u & u'\omega \\ u'\omega^* & u \end{pmatrix} e^{i\xi (\mathbf{G}_1^m + \mathbf{G}_2^m) \cdot \mathbf{r}}$$

Here,  $u = 0.0797$  eV and  $u' = 0.0975$  eV are coupling constants<sup>49</sup> that give the strength of the interaction between like ( $A \leftrightarrow A$ ,  $B \leftrightarrow B$ ) and opposing ( $A \leftrightarrow B$ ) sublattices in the two layers, the difference of which accounts for out-of-plane corrugation,  $\omega = e^{2\pi i/3}$ , and  $U$  is the Hermitian conjugate of  $U$ . The moiré reciprocal lattice vectors,  $\mathbf{G}_j^m = \mathbf{a}_j^{(1)} - \mathbf{a}_j^{(2)}$ , are given by the difference between the reciprocal lattice vectors in the upper ( $\mathbf{a}_j^{(1)}$ ) and lower ( $\mathbf{a}_j^{(2)}$ ) layers.

Magnetic field effects can be included by making the substitution  $\mathbf{k} \rightarrow \mathbf{k} + e\mathbf{A}/\hbar$  in the effective Hamiltonian. Here,  $\mathbf{A}$  is the vector potential, which is related to the static magnetic field via  $\mathbf{B} = \nabla \times \mathbf{A}$ . In general, the bandstructure in a magnetic field cannot be computed because the addition of a spatially dependent vector potential breaks translational invariance. However, at certain values of the magnetic field—specifically when  $SB/(h/e) = p/q$ , where  $p$  and  $q$  are co-prime integers and  $S$  is the area of the unit cell—a ‘magnetic’ unit cell can be introduced

whereupon it becomes possible to solve the Schrödinger equation using the corresponding ‘magnetic’ Bloch conditions<sup>50</sup>. It is then possible to construct a Hamiltonian matrix in the basis of the monolayer graphene Landau levels<sup>51,52</sup>. Although the Landau levels basis is unbounded, the Hamiltonian matrix can be truncated at an energy where the higher-energy Landau levels only weakly affect the low-energy spectrum. This cut-off energy must be much larger than the interlayer coupling characterized by the coupling constants  $u$  and  $u'$ . The resulting finite matrix can then be diagonalized. This results in a bandstructure diagram in terms of  $p/q$  that is directly related to the strength of the magnetic field and indirectly related to the twist angle, because the moiré unit cell area,  $S = \sqrt{3} a^2 / [8 \sin^2(\theta/2)]$ , depends on  $\theta$ . For a varying magnetic field or twist angle the bands are computed for each individual parameter value assuming that these values are homogeneous throughout the material.

Level crossings in the bandstructure are observed as the magnetic field or the twist angle are varied. These occur as a result of the ‘Rashba-like’ splitting of the dispersive bands. In general, this type of splitting leads to two Landau level series, largely overlapping in energy, which cross as a function of magnetic field<sup>53</sup> (Extended Data Fig. 10a), and as a function of  $\theta$  (Extended Data Fig. 10b), as is the case in the experimental data. These are due to the evolution of the Rashba-like splitting with  $\theta$  (Extended Data Fig. 10c–e).

## Origin of the internal in-plane electric field and numerical electrostatic simulations

The origin of the internal electric field in presence of twist-angle gradients can be understood intuitively as follows. An external in-plane electric field applied to graphene—for example, by charge disorder in the substrate—exerts a force on the electrons. Thermal equilibrium conditions require zero net force on the carriers, and so charge redistribution will occur, creating an opposing electric field and leading to screening of the external field, thus achieving the required zero net force. This is the common situation in metals. In the case of MATBG, the backgate voltage induces a nearly uniform carrier density, while the twist-angle disorder induces a variable chemical potential  $\mu(\mathbf{r})$  (Fig. 4b), which exerts an in-plane force on the carriers,  $\mathbf{F}_{\parallel} = -\nabla\mu$ . To attain zero net force in thermal equilibrium, an in-plane electric field  $\mathbf{E}_{\parallel} = \nabla\mu/q$  must therefore be generated by carrier redistribution. Thus, in contrast to common charge disorder, in which the system tends to screen external electric fields, in the case of twist-angle disorder the system counterintuitively spontaneously generates internal electric fields to counterbalance the force produced by the variable chemical potential. These internal fields do not generate current at zero magnetic field; however, in a finite magnetic field transverse topological and nontopological currents will be induced in the ground state.

For the results presented in Fig. 4, simulations (using the COMSOL Multiphysics analysis software) were used to solve electrostatic equations for the potential  $V$  and charge density  $\rho = -en_c$  at  $B_a = 0$  and in the quantum Hall state at  $B_a = 1.22$  T. The simulations included a backgate electrode at a constant electric potential  $V_{bg}$  and a grounded MATBG in a  $3 \times 0.5 \mu\text{m}^2$   $x$ - $z$  box, assuming translation invariance along the  $y$  axis, with boundary conditions of  $E_{\perp} = 0$  on the external surfaces of the box. An iterative self-consistent solution for  $V(x, y, z)$  and  $\rho(x, y)$  was obtained, satisfying the following conditions: (1)  $V$  depends on  $\rho$  through  $\nabla \cdot \mathbf{E} = \rho/\epsilon_r \epsilon_0$  and  $\mathbf{E} = -\nabla V$ , where  $\epsilon_r$  is the relative permittivity (we took  $\epsilon_r = 4$  for hBN) and  $\epsilon_0$  is the vacuum permittivity (here  $\nabla = (\partial_x, \partial_y, \partial_z)$  is the three dimensional operator); and (2)  $\rho$  depends on  $V$  through the integrated DOS  $\mathcal{N}_c(\mu; \theta)$ , where  $\mu(x, y) = -qV(x, y)$  and  $q = \pm e$  is the carrier charge (negative sign for  $\mu > 0$ ). The integrated DOS  $\mathcal{N}_c(\mu; \theta)$  was calculated for  $B = 0$  and  $B \neq 0$  as described in the Methods section ‘Bandstructure calculations and Landau level crossings’.

Once  $V(x, y, z = 0)$  and  $\rho(x, y)$  were found in the plane of the MATBG, the incompressible surface currents were calculated using  $\mathbf{J}^{\parallel} = -\sigma \nabla V$ ,

where  $\sigma_{xy}(x, y) = -\sigma_{yx}(x, y) = -v(x, y)e^2/h$ , and  $\sigma_{xx} = \sigma_{yy} = 0$  are the components of the conductivity tensor  $\sigma$ .

## Measurement parameters

All the measurements were carried out at  $T = 300$  mK in an out-of-plane applied magnetic field  $B_a$ .

**Figure 1b and Extended Data Fig. 2b.** Device B,  $I_{ac} = 10$  nA (r.m.s.),  $V_{bg} = -15$  V to 15 V.

**Figure 1d.** Device B,  $B_a = 1.08$  T, SOT diameter 250 nm, scan height 40 nm,  $V_{bg}^{ac} = 20$  mV (r.m.s.),  $\Delta V_{bg} = 6.25$  mV, acquisition time 6 s per point, total acquisition time 8 h.

**Figure 1f and Extended Data Figs. 6a, 7.** Device A,  $B_a = 1.19$  T, SOT diameter 220 nm, scan height 100 nm,  $V_{bg}^{ac} = 15$  mV (r.m.s.),  $V_{bg} = 4.7$  mV, acquisition time 6 s per point, total acquisition time 12 h.

**Figure 2.** Device A,  $B_a = 1.22$  T, SOT diameter 220 nm, scan height 60 nm,  $V_{bg}^{ac} = 35$  mV (r.m.s.), pixel size 26 nm, 160 ms per pixel, total acquisition time 21.4 h.

**Figure 3a and Supplementary Video 2.** Device A,  $B_a = 1.16$  T, SOT diameter 220 nm, scan height 110 nm,  $V_{bg}^{ac} = 80$  mV (r.m.s.), pixel size 60 nm, 60 ms per pixel, acquisition time 60 min per frame.

**Supplementary Video 1 and Extended Data Figs. 6b, 8a, b.** Device A,  $B_a = 1.22$  T, SOT diameter 220 nm, scan height 60 nm,  $V_{bg}^{ac} = 35$  mV (r.m.s.), pixel size 57 nm, 60 ms per pixel, acquisition time 30 min per frame.

**Figure 3e.** Device B,  $B_a = 1.08$  T, SOT diameter 250 nm, scan height 140 nm,  $V_{bg}^{ac} = 60$  mV (r.m.s.), pixel size 50 nm, 60 ms per pixel, acquisition time 33 min.

**Supplementary Videos 3, 4 and Extended Data Fig. 8c, d.** Device B,  $B_a = 1.08$  T, SOT diameter 250 nm, scan height 70 nm (Video 3) and 80 nm (Video 4),  $V_{bg}^{ac} = 60$  mV (r.m.s.), pixel size 43 nm, 60 ms per pixel, acquisition time 25 min per frame.

**Extended Data Fig. 2a.** Device A,  $I_{ac} = 10$  nA (r.m.s.),  $V_{bg} = -17$  V to 17 V, Si backgate 50 V.

**Extended Data Fig. 3a.** Device A,  $I_{ac} = 5$  nA (r.m.s.),  $V_{bg} = -6$  V to  $-3.5$  V, Si backgate 50 V.

**Extended Data Fig. 3b.** Device B,  $I_{ac} = 4$  nA (r.m.s.),  $V_{bg} = -7$  V to  $-4$  V.

**Extended Data Fig. 3c.** Device B,  $I_{ac} = 10$  nA (r.m.s.),  $V_{bg} = -6.5$  V to  $-4.7$  V,  $B_a = 0$  T.

## Data availability

The data that support the findings of this study are available from the corresponding authors on reasonable request.

- Kim, K. et al. Tunable moiré bands and strong correlations in small-twist-angle bilayer graphene. *Proc. Natl Acad. Sci. USA* **114**, 3364–3369 (2017).
- Wang, L. et al. One-dimensional electrical contact to a two-dimensional material. *Science* **342**, 614–617 (2013).
- Anahory, Y. et al. SQUID-on-tip with single-electron spin sensitivity for high-field and ultra-low temperature nanomagnetic imaging. *Nanoscale* **12**, 3174–3182 (2020).
- Huber, M. E. et al. DC SQUID series array amplifiers with 120 MHz bandwidth. *IEEE Trans. Appl. Supercond.* **11**, 1251–1256 (2001).
- Finkler, A. et al. Scanning superconducting quantum interference device on a tip for magnetic imaging of nanoscale phenomena. *Rev. Sci. Instrum.* **83**, 073702 (2012).
- Finkler, A. et al. Self-aligned nanoscale SQUID on a tip. *Nano Lett.* **10**, 1046–1049 (2010).
- Lachman, E. O. et al. Visualization of superparamagnetic dynamics in magnetic topological insulators. *Sci. Adv.* **1**, e1500740 (2015).

40. Halbertal, D. et al. Nanoscale thermal imaging of dissipation in quantum systems. *Nature* **539**, 407–410 (2016).
41. Kleinbaum, E. & Csáthy, G. A. Note: a transimpedance amplifier for remotely located quartz tuning forks. *Rev. Sci. Instrum.* **83**, 126101 (2012).
42. Geller, M. R. & Vignale, G. Currents in the compressible and incompressible regions of the two-dimensional electron gas. *Phys. Rev. B* **50**, 11714–11722 (1994).
43. Kim, P. Graphene and relativistic quantum physics. In *Dirac Matter* (eds Duplantier B., Rivasseau V. & Fuchs J. N.) 1–23 (Birkhäuser, 2017).
44. Dean, C. R. et al. Boron nitride substrates for high-quality graphene electronics. *Nat. Nanotechnol.* **5**, 722–726 (2010).
45. Martin, J. et al. Observation of electron–hole puddles in graphene using a scanning single-electron transistor. *Nat. Phys.* **4**, 144–148 (2008).
46. Lopes dos Santos, J. M. B., Peres, N. M. R. & Castro Neto, A. H. Graphene bilayer with a twist: electronic structure. *Phys. Rev. Lett.* **99**, 256802 (2007).
47. Kindermann, M. & First, P. N. Local sublattice-symmetry breaking in rotationally faulted multilayer graphene. *Phys. Rev. B* **83**, 045425 (2011).
48. Koshino, M. & Moon, P. Electronic properties of incommensurate atomic layers. *J. Phys. Soc. Jpn.* **84**, 121001 (2015).
49. Koshino, M. et al. Maximally localized Wannier orbitals and the extended Hubbard model for twisted bilayer graphene. *Phys. Rev. X* **8**, 031087 (2018).
50. Xiao, D., Chang, M.-C. & Niu, Q. Berry phase effects on electronic properties. *Rev. Mod. Phys.* **82**, 1959–2007 (2010).
51. Bistritzer, R. & MacDonald, A. H. Moiré butterflies in twisted bilayer graphene. *Phys. Rev. B* **84**, 035440 (2011).
52. Moon, P. & Koshino, M. Energy spectrum and quantum Hall effect in twisted bilayer graphene. *Phys. Rev. B* **85**, 195458 (2012).
53. Mireles, F. & Schliemann, J. Energy spectrum and Landau levels in bilayer graphene with spin–orbit interaction. *New J. Phys.* **14**, 093026 (2012).

**Acknowledgements** We thank A. Stern and E. Berg for valuable discussions and M. F. da Silva for constructing the COMSOL simulations. This work was supported by the Sagol WIS–MIT Bridge Program, by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant no. 785971), by the Israel Science Foundation (ISF, grant no. 994/19), by the Minerva Foundation with funding from the Federal German Ministry of Education and Research, and by the Leona M. and Harry B. Helmsley Charitable

Trust grant no. 2018PG-ISL006. Y.C., P.J.-H. and E.Z. acknowledge the support of the MISTI (MIT International Science and Technology Initiatives) MIT–Israel Seed Fund. Work at MIT was supported by the National Science Foundation (NSF, grant no. DMR-1809802), the Center for Integrated Quantum Materials under NSF grant no. DMR-1231319 and the Gordon and Betty Moore Foundation's EPIQS Initiative through grant no. GBMF4541 to P.J.-H. for device fabrication, transport measurements and data analysis. This work was performed in part at the Harvard University Center for Nanoscale Systems (CNS), a member of the National Nanotechnology Coordinated Infrastructure Network (NNCI), which is supported by the National Science Foundation under NSF ECCS award no. 1541959. D.R.-L. acknowledges partial support from Fundació Bancaria 'la Caixa' (LCF/BQ/AN15/1038001) and from the US Army Research Office grant no. W911NF-17-S-0001. M.K. acknowledges the financial support of JSPS KAKENHI grant no. JP17K05496. J.A.C. and P.M. were supported by the Science and Technology Commission of Shanghai Municipality grant no. 19ZR1436400, the NYU–ECNU Institute of Physics at NYU Shanghai and New York University Global Seed Grants for Collaborative Research. J.A.C. acknowledges support from the National Science Foundation of China grant no. 11750110420. This research was carried out on the High Performance Computing resources at NYU Shanghai. K.W. and T.T. acknowledge support from the Elemental Strategy Initiative conducted by the MEXT, Japan, A3 Foresight by JSPS and the CREST (JPMJCR15F3), JST.

**Author contributions** A.U., S.G. and E.Z. designed the experiment. A.U., S.G. and Y.C. performed the measurements. A.U. and S.G. performed the analysis. Y.C., D.R.-L. and P.J.-H. designed and provided the samples and contributed to the analyses of the results. K.B. fabricated the SOTs. Y.M. fabricated the tuning forks. J.A.C. performed the tight-binding calculations with P.M. and M.K., and K.W. and T.T. fabricated the hBN. A.U., S.G. and E.Z. wrote the manuscript. All authors participated in discussions and in writing of the manuscript.

**Competing interests** The authors declare no competing interests.

#### Additional information

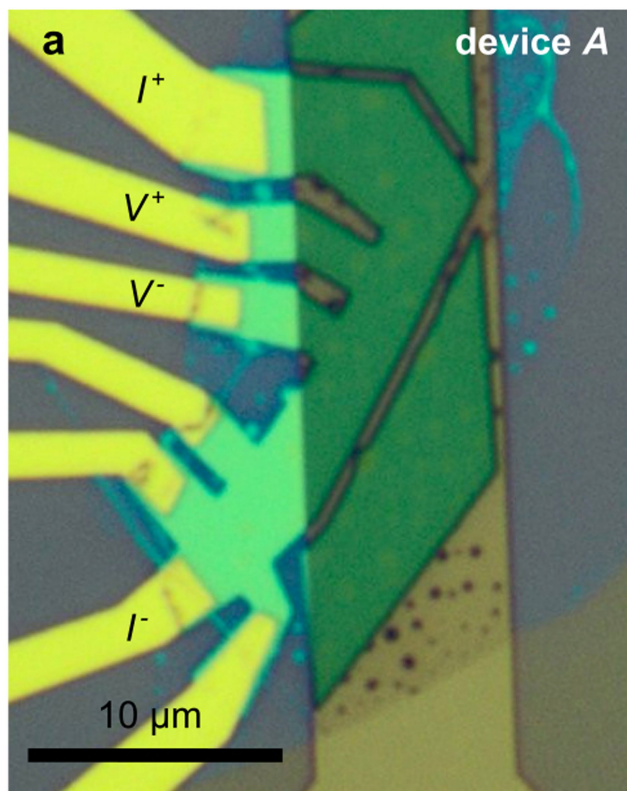
**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41586-020-2255-3>.

**Correspondence and requests for materials** should be addressed to P.J.-H. or E.Z.

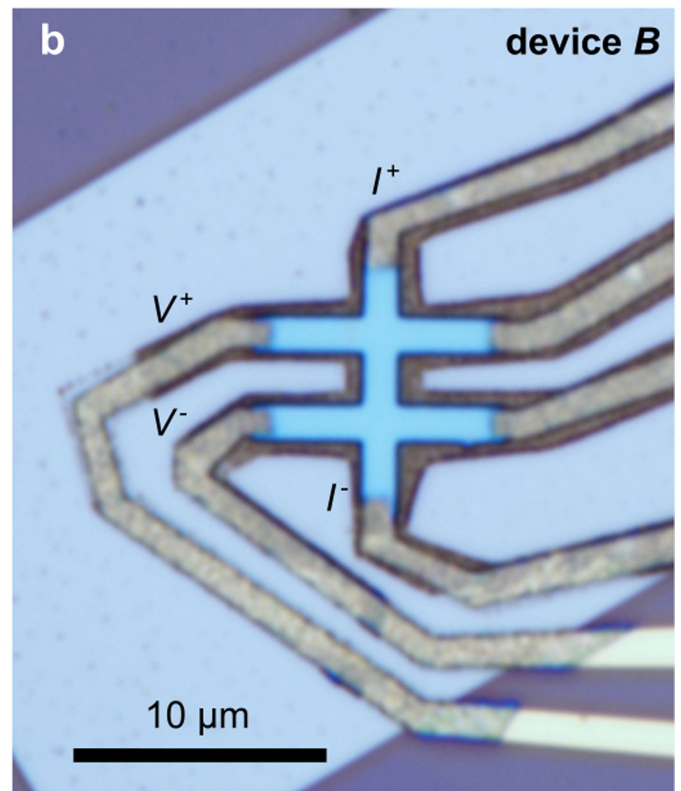
**Peer review information** *Nature* thanks Emanuel Tutuc, Fan Zhang and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>.

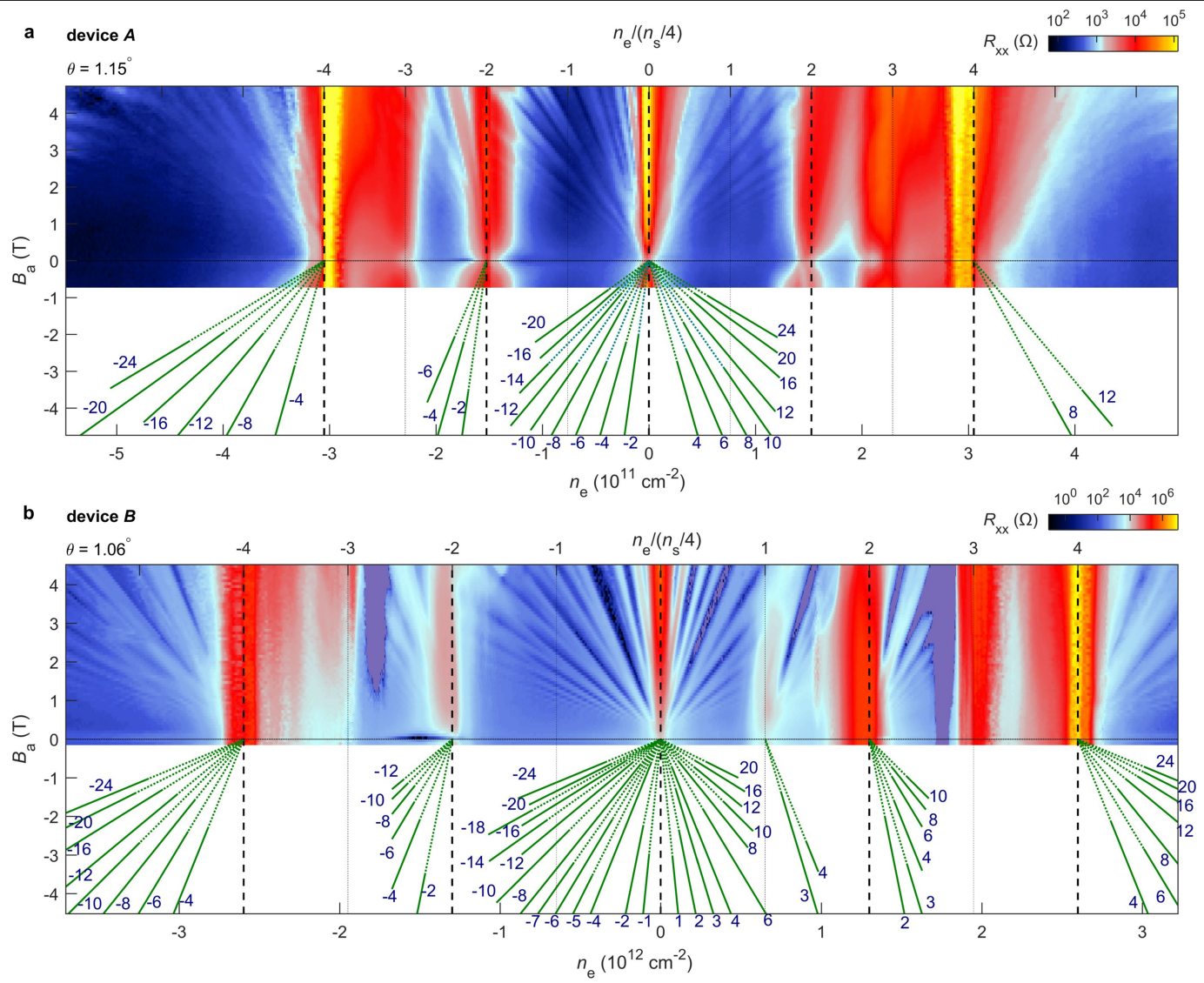




**Extended Data Fig. 1 | Optical image of MATBG devices.** **a**, Optical image of device A showing hBN/MATBG/hBN (green), the underlying PdAu backgate (light brown) and the marked electrodes used for four-probe  $R_{xx}$

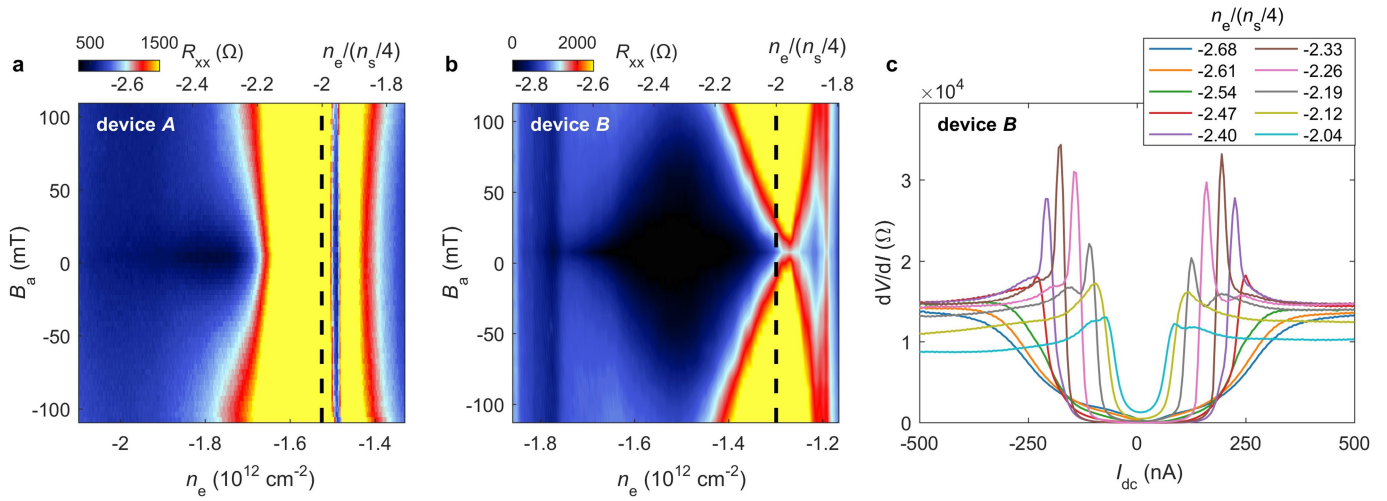


measurements. **b**, Optical image of device B (cyan) on the PdAu backgate (light blue) with marked electrodes.



**Extended Data Fig. 2 | Transport measurements at  $T=300$  mK. a,** Four-probe measurement of  $R_{xx}(V_{bg})$  versus  $B_a$  in device A using an excitation current of 10 nA with the corresponding traces of the Landau fan diagram at the bottom.

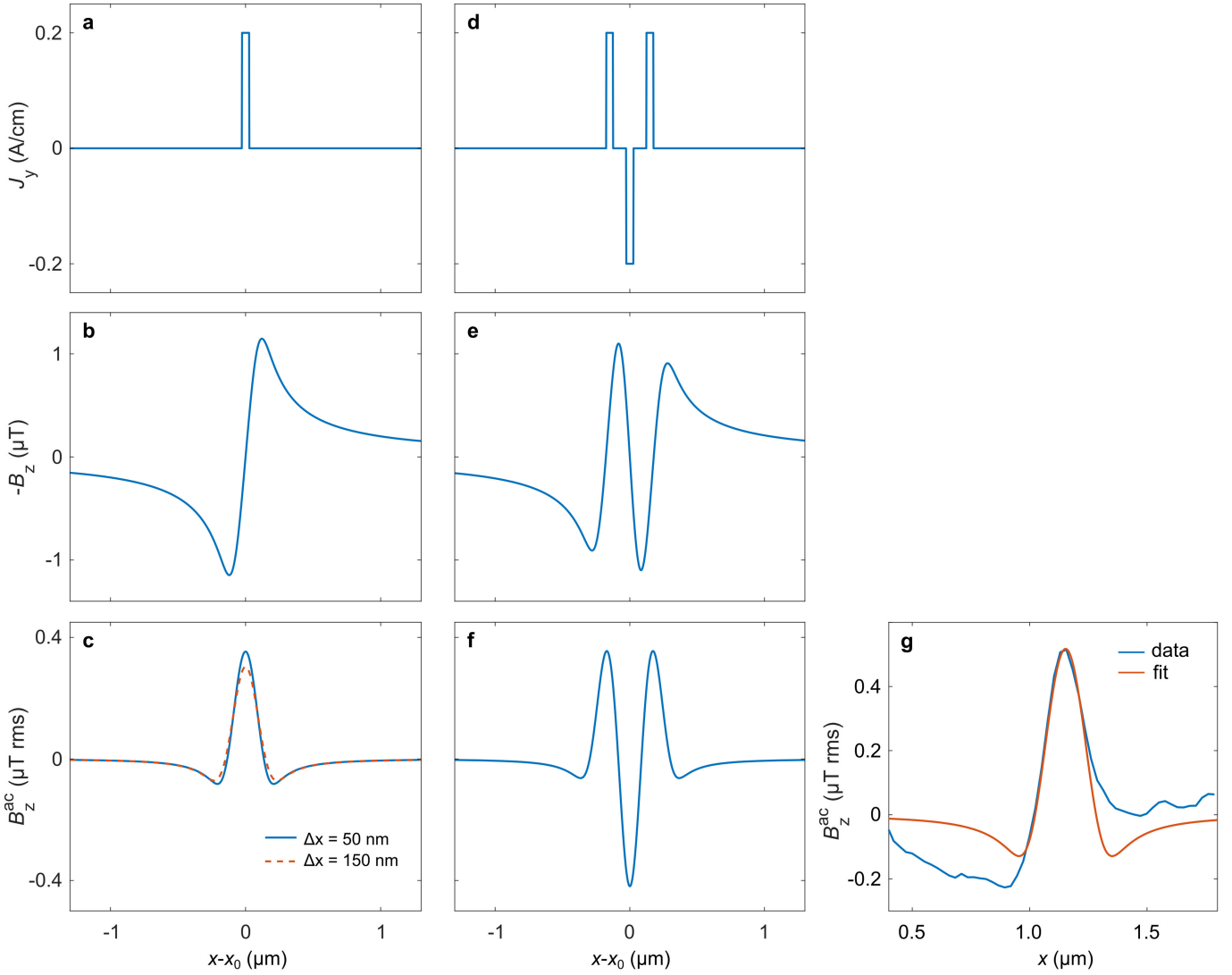
The green solid lines show the segments that can be traced in the data and the dotted lines indicate their extrapolation to the origin. **b,** As in **a** for device B. The purple colour marks the regions where the  $R_{xx}$  signal was slightly negative.



**Extended Data Fig. 3 | Transport measurements in the superconducting state at  $T=300$  mK. a, b,** Colour rendering of  $R_{xx}$  measured in the vicinity of  $-n_s/2$  versus  $B_a$  and  $n_e$  at low fields using an r.m.s. excitation current of 5 nA in device A (a) and 4 nA in device B (b). A zero-resistance superconducting state

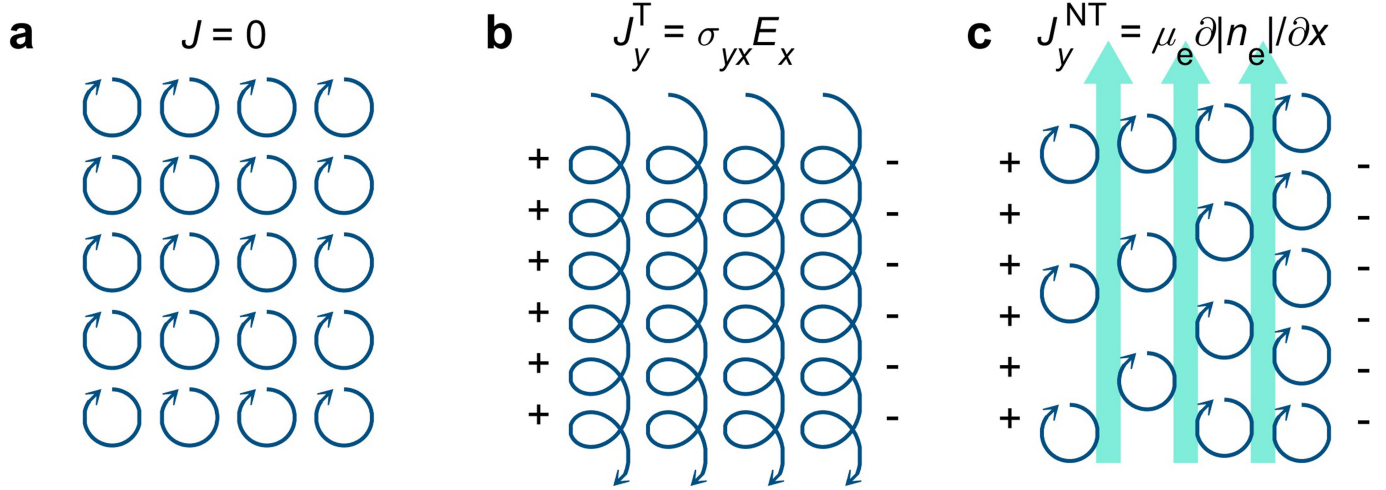
(black) is observed in device B. c,  $dV/dI$  versus  $I_{dc}$  characteristics at various carrier concentrations  $n_e$  in the superconducting state in device B at  $B_a=0$  T using an r.m.s. a.c. excitation  $I_{ac}=10$  nA.





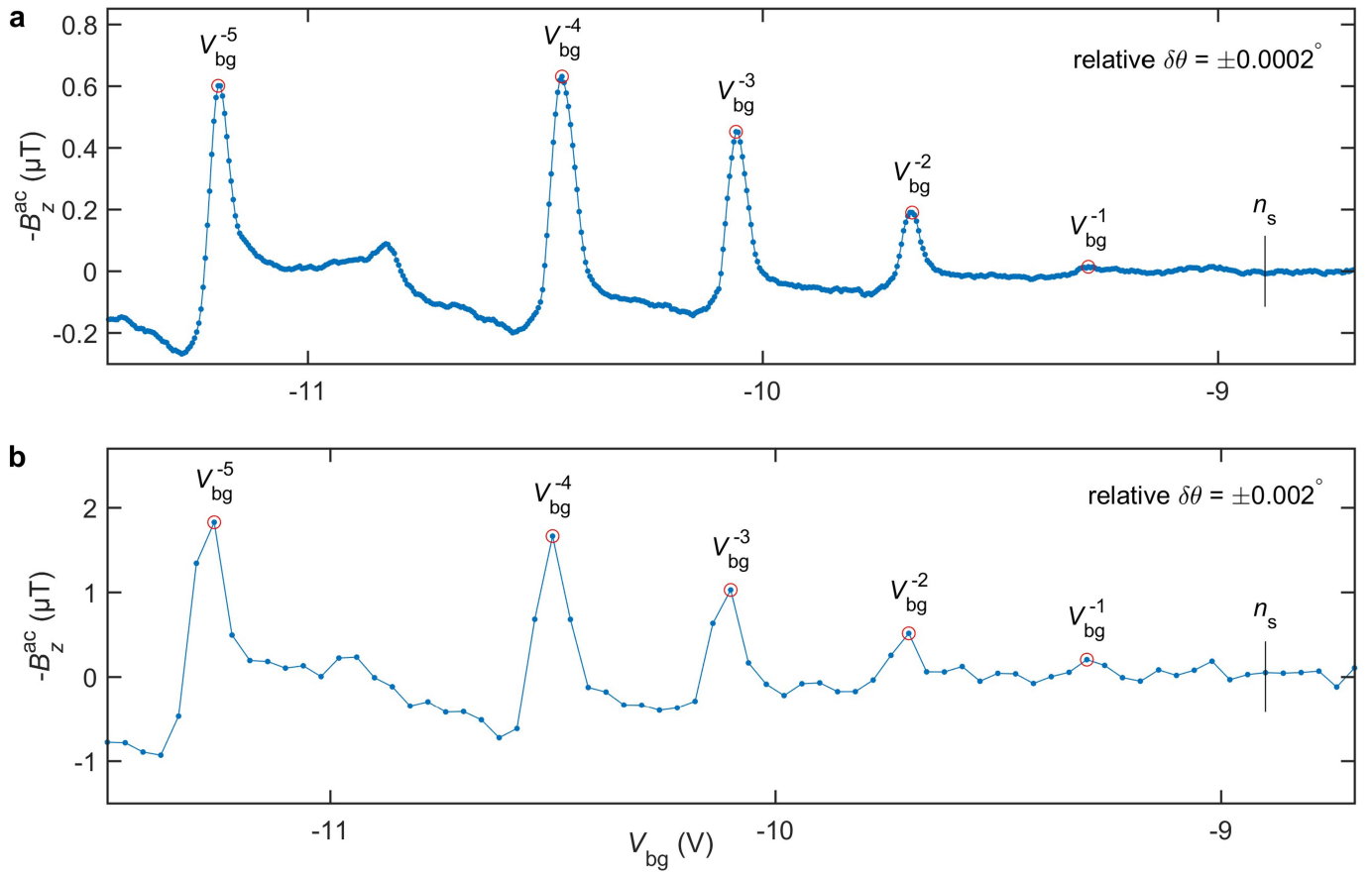
**Extended Data Fig. 4 | Numerical simulation demonstrating current imaging by measuring  $B_z^{\text{ac}}$ .** **a**, Current distribution  $J_y(x-x_0)$  of a  $\Delta x = 50$ -nm-wide channel carrying  $I_y = 1 \mu\text{A}$  in the  $\hat{y}$  direction. **b**, Calculated  $-B_z(x-x_0)$  at a height of 70 nm above the sample, convoluted with a 220-nm-diameter SOT sensing area. **c**, Calculated  $B_z^{\text{ac}}(x-x_0)$  for an r.m.s.  $x_0^{\text{ac}} = 54$  nm spatial modulation of the channel position. The dashed profile corresponds to a current strip of width  $\Delta x = 150$  nm carrying the same current, showing that the spatial resolution is limited by the SOT diameter. **d–f**, As in **a–c** but for three counter-propagating currents spaced 150 nm apart. **g**, Analysis of the  $B_z^{\text{ac}}$  peak of an incompressible strip.  $B_z^{\text{ac}}(x)$  signal (blue) acquired along the

line indicated in Fig. 3a for  $V_{\text{bg}} = -10.54$  V (a single vertical line from Fig. 2a) showing the  $v = -12$  incompressible peak, along with a numerical fit (red). The fit uses the experimental values of  $V_{\text{bg}}^{\text{ac}}$ ,  $h_{\text{SOT}}$  and the SOT diameter with a single fitting parameter of the total current in the incompressible strip resulting in  $I^{\text{I}} = 1.3 \mu\text{A}$ . An incompressible strip of width  $\Delta x = 50$  nm was used for the fit. The mean value of  $B_z^{\text{ac}}(x)$  was subtracted from the data. The asymmetry in  $B_z^{\text{ac}}(x)$  away from the peak is caused by the presence of counterflowing nontopological currents  $I^{\text{NT}}$  of lower density in the adjacent compressible strips.



**Extended Data Fig. 5 | The origin of equilibrium currents in the compressible and incompressible quantum Hall states. a,** Semiclassical picture of cyclotron orbits of holes with mutually canceling neighbouring currents, resulting in zero bulk current. **b,** In the presence of an in-plane electric field  $E_x$  (+ and - signs represent external charges) the cyclotron orbits

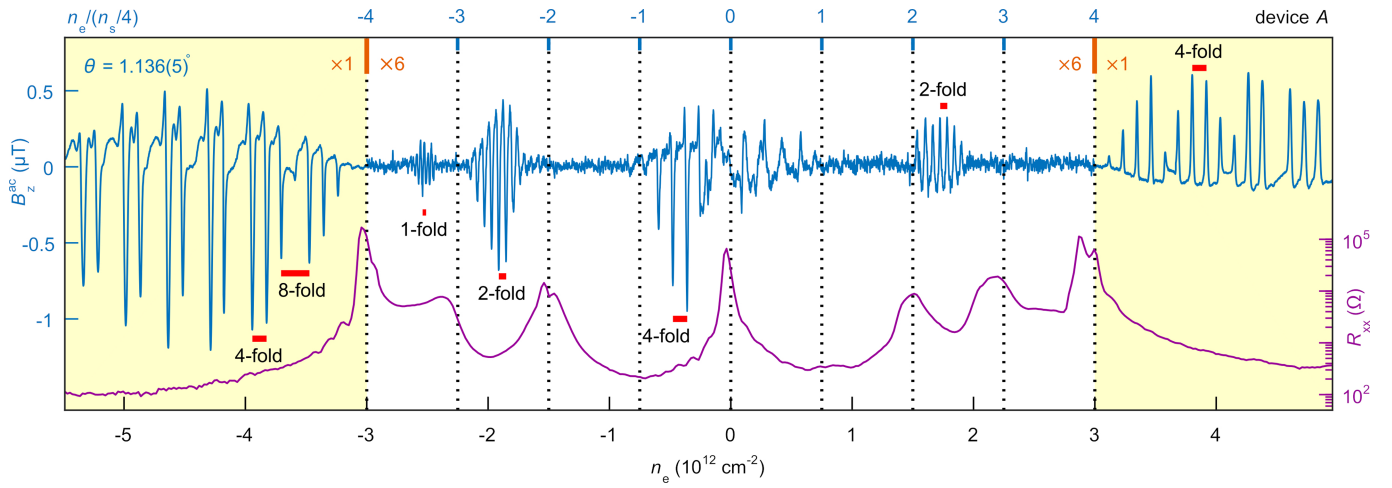
acquire a drift velocity resulting in a non-zero  $J_y^T$  in the incompressible state. **c,** In the compressible regime the external in-plane electric field is screened by establishing a charge-density gradient, giving rise to  $J_y^{NT}$  flowing in the opposite direction (cyan arrows).



**Extended Data Fig. 6 | Determination of the accuracy of the twist-angle measurement. a**, Traces of  $-B_z^{\text{ac}}$  versus  $V_{\text{bg}}$  in device A (from Fig. 1f) acquired with a step size  $\Delta V_{\text{bg}} = 4.7$  mV and an r.m.s.  $V_{\text{bg}}^{\text{ac}} = 15$  mV. The positions of the  $V_{\text{bg}}^{N=-3}$  and  $V_{\text{bg}}^{N=-4}$  peaks can be determined to an accuracy better than  $\pm \Delta V_{\text{bg}}$  (one step size), corresponding to a relative  $\theta$  accuracy of  $\delta\theta = \pm 0.0002^\circ$ . **b**, As in **a**, taken

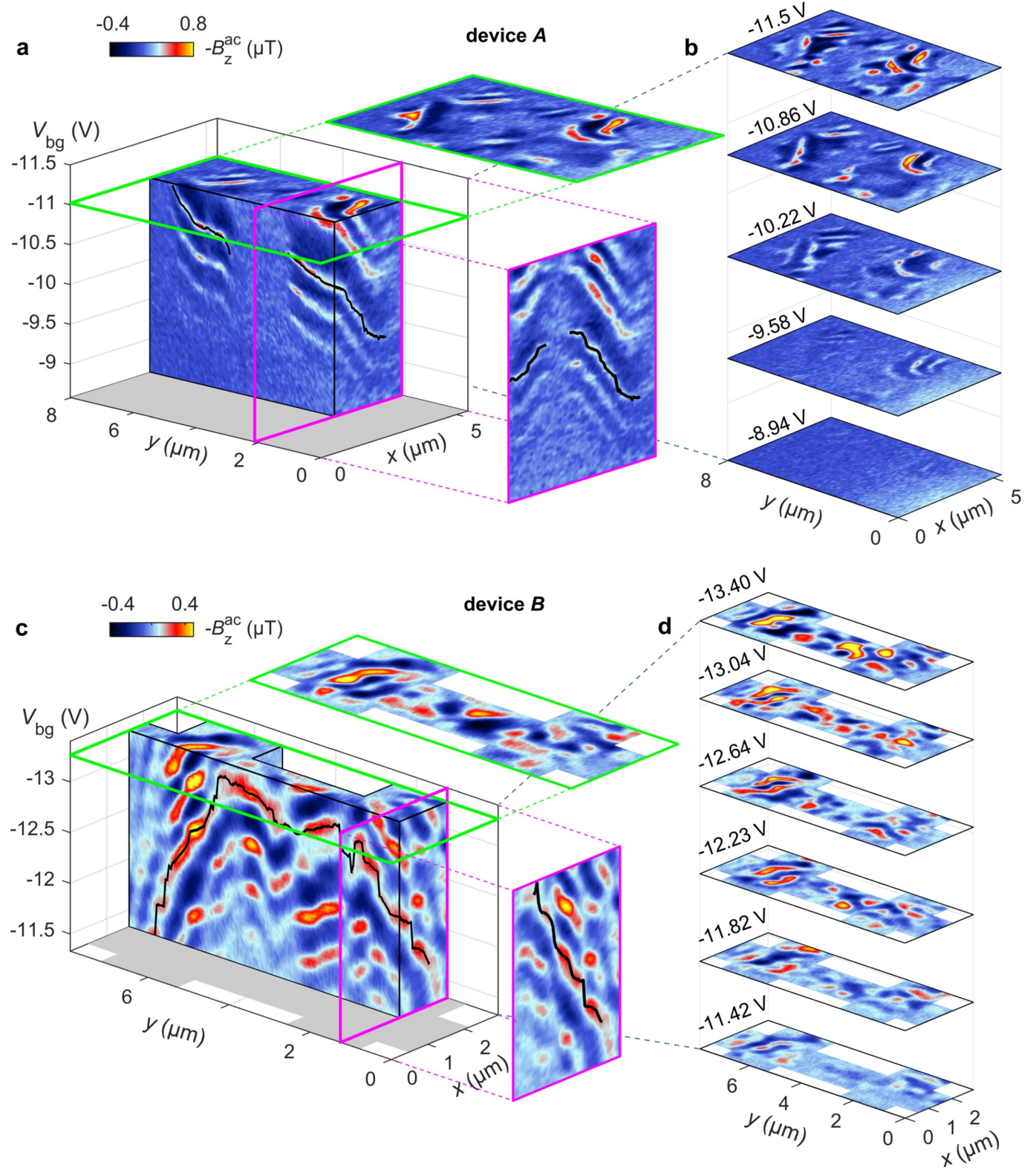
from Supplementary Video 1 at a pixel position  $(x, y) = (2.53 \mu\text{m}, 5.9 \mu\text{m})$  with step size  $\Delta V_{\text{bg}} = 40$  mV and an r.m.s.  $V_{\text{bg}}^{\text{ac}} = 35$  mV, resulting in a relative  $\theta$  accuracy of  $\delta\theta = \pm 0.002^\circ$  in the imaging mode. The larger  $B_z^{\text{ac}}$  signal and the broader  $I^{\text{r}}$  peaks in **b** compared to **a** are due to larger  $V_{\text{bg}}^{\text{ac}}$  excitation (see Methods section 'Measurement parameters').





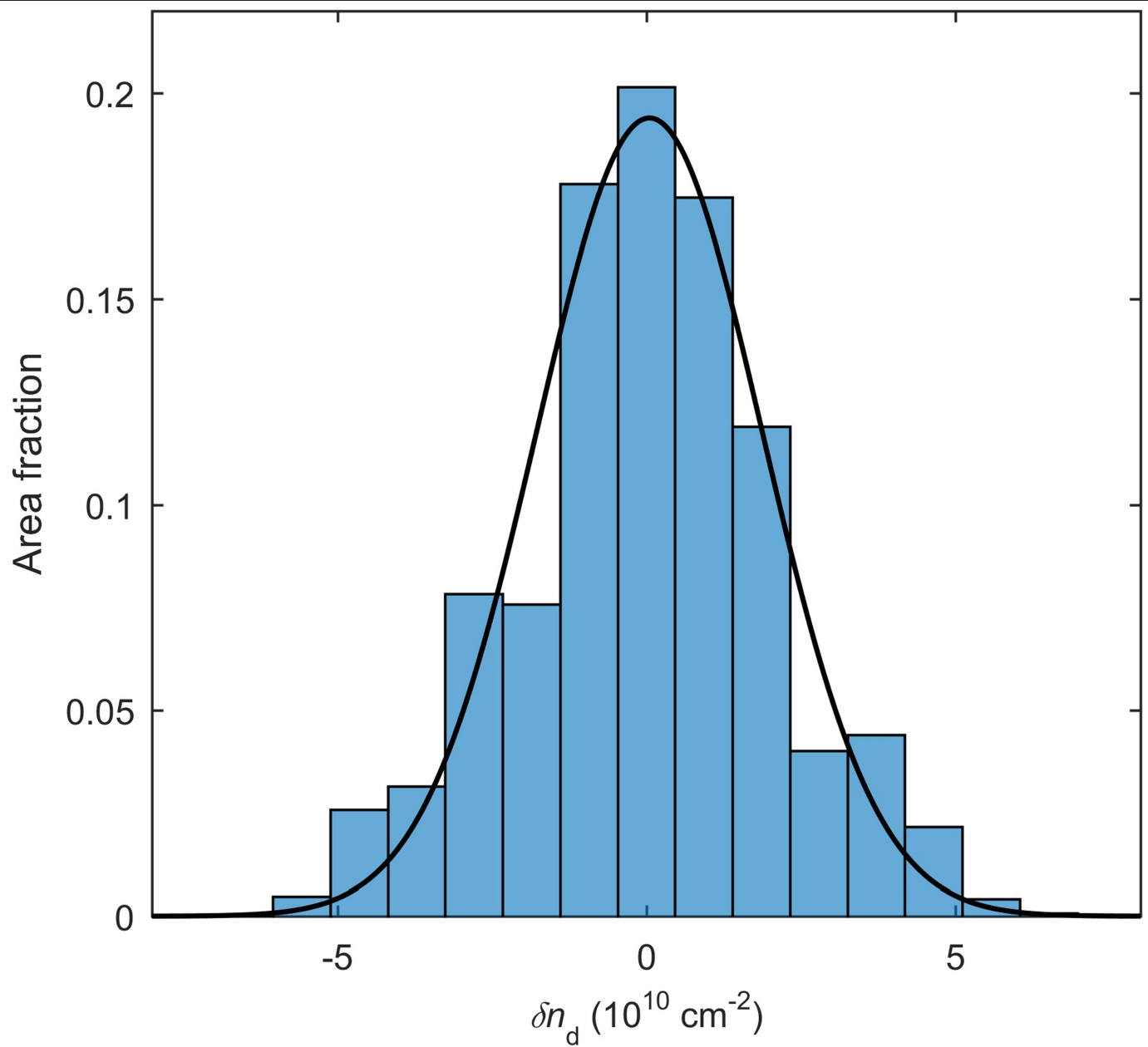
**Extended Data Fig. 7 | Resolving the local quantum Hall states in flat and dispersive bands in device A.** Global  $R_{xx}$  (purple, right axis) and local  $B_z^{\text{ac}}$  (blue, left axis) measured at a point in the bulk of device A versus the electron density  $n_e$  at  $B_a = 1.19 \text{ T}$ . The sharp  $B_z^{\text{ac}}$  peaks reflect the  $I^T$  current in

incompressible strips with sign determined by the sign of  $\sigma_{yy}$ , magnitude by the Landau level energy gap and separation by the Landau level degeneracy (red bars). The dispersive bands are shaded in yellow and the signal in the flat bands is amplified six times for clarity.



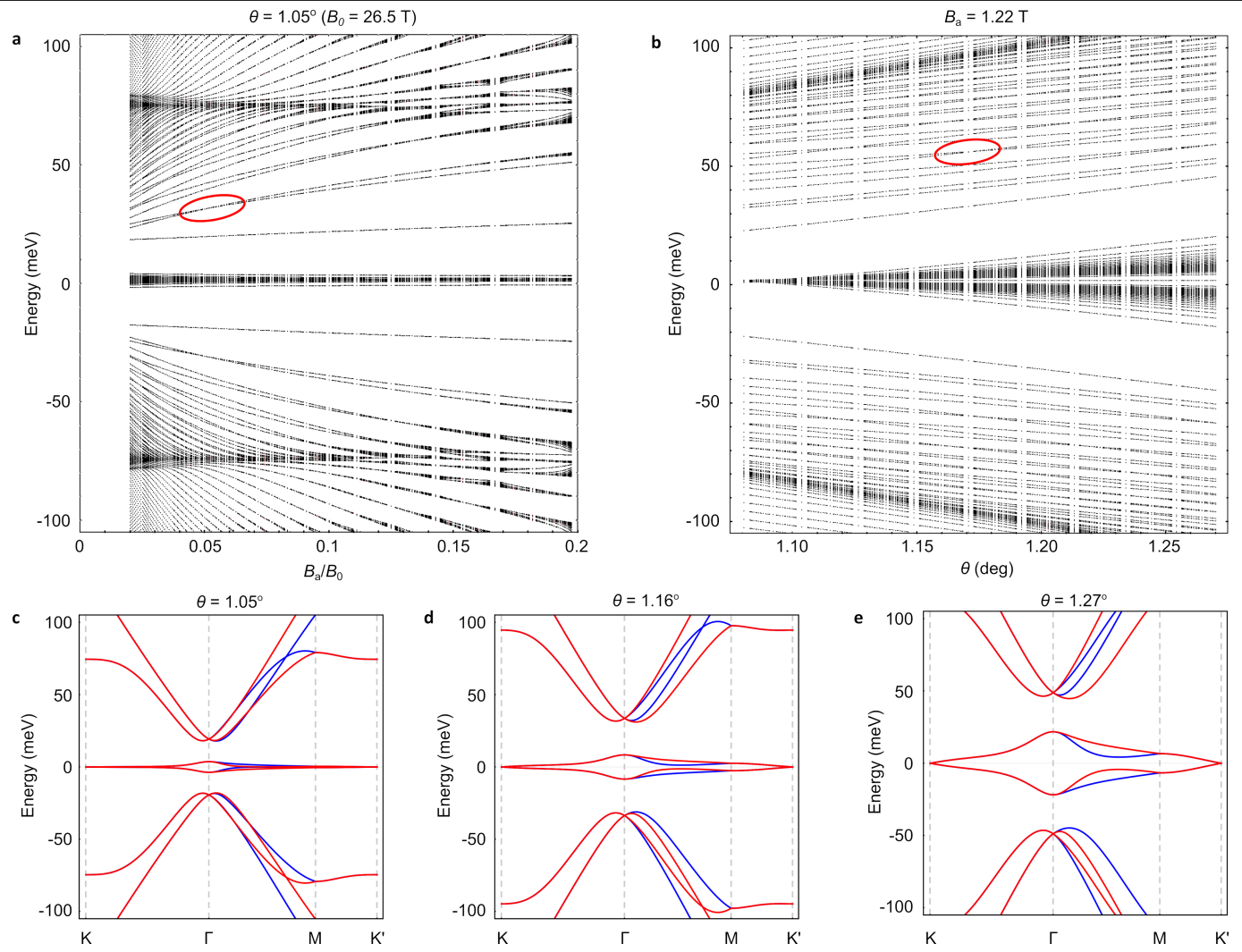
**Extended Data Fig. 8 | Landau level tomography.** **a**, Slices of the 3D dataset  $B_z^{\text{ac}}(x, y, V_{\text{bg}})$  along various planes for device A. The bright signals denote the 2D manifolds tracing the incompressible states. The black lines trace the  $N = -4$  incompressible manifold used to determine  $n_s(x, y)$  and  $\theta(x, y)$ . It separates fourfold degenerate Landau levels below it from an eightfold degenerate Landau level above it (wide dark blue band). The region in the centre of the sample that shows no Landau levels corresponds to the grey-blue area in Fig. 3b where no MATBG physics is resolved. **b**, Representative horizontal slices of the

data from Supplementary Video 1 showing the evolution of the Landau levels with  $V_{\text{bg}}$ . **c**, As in **a**, for device B. For the range of gate voltages shown,  $\varepsilon_f$  lies in the p dispersive band for the entire sample. The black lines show an example of a trace of the incompressible manifold lying above an eightfold degenerate Landau level. **d**, Representative horizontal slices of the data from Supplementary Video 3. An interactive interface for tomographic visualization of the data is available at ref. <sup>30</sup>.



**Extended Data Fig. 9 | Histogram of the charge disorder in device B.** Histogram of  $\delta n_d(r)$  data from Fig. 3h along with a Gaussian fit (black) with a standard deviation  $\Delta n_d = 2.59 \times 10^{10} \text{ cm}^{-2}$ .





**Extended Data Fig. 10 | Landau level crossings of the dispersive bands.**

**a**, Numerically calculated Landau level energies as a function of magnetic field for a fixed  $\theta = 1.05^\circ$ . An example level crossing is highlighted in red.

**b**, Numerically calculated Landau level energies as a function of  $\theta$  for a fixed

$B_a = 1.22$  T. An example level crossing is highlighted in red. **c–e**, The  $B_a = 0$  bandstructure of bilayer graphene for  $\theta = 1.05^\circ$  (**c**),  $\theta = 1.16^\circ$  (**d**) and  $\theta = 1.27^\circ$  (**e**). The blue and red lines indicate the bands that arise from the positive and negative valleys, respectively.

# Iron-based binary ferromagnets for transverse thermoelectric conversion

<https://doi.org/10.1038/s41586-020-2230-z>

Received: 26 July 2019

Accepted: 4 February 2020

Published online: 27 April 2020

 Check for updates

Akito Sakai<sup>1,2,3,10</sup>, Susumu Minami<sup>4,5,10</sup>, Takashi Koretsune<sup>6,10</sup>, Taishi Chen<sup>1,3,10</sup>, Tomoya Higo<sup>1,3,10</sup>, Yangming Wang<sup>1</sup>, Takuya Nomoto<sup>7</sup>, Motoaki Hirayama<sup>5</sup>, Shinji Miwa<sup>1,3,8</sup>, Daisuke Nishio-Hamane<sup>1</sup>, Fumiyuki Ishii<sup>4,5</sup>, Ryotaro Arita<sup>3,5,7</sup> & Satoru Nakatsuji<sup>1,2,3,8,9</sup>✉

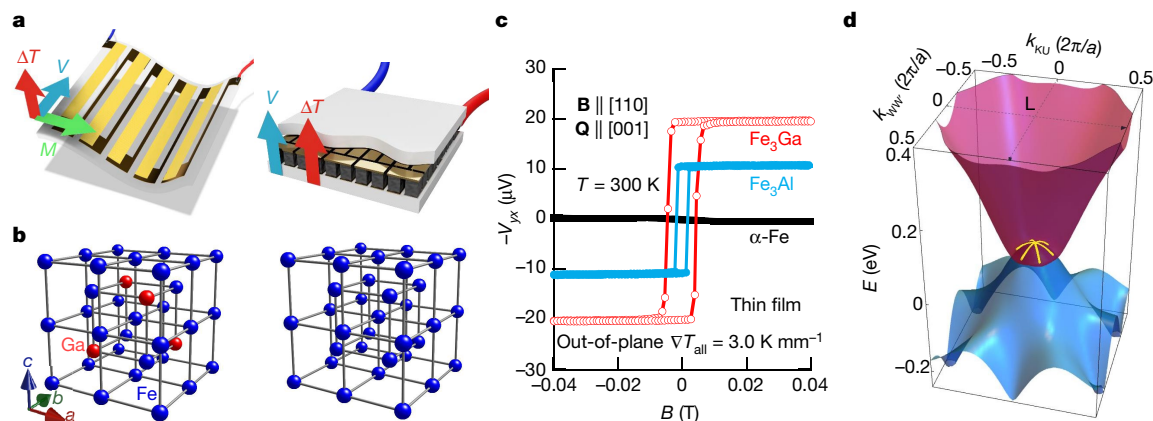
Thermoelectric generation using the anomalous Nernst effect (ANE) has great potential for application in energy harvesting technology because the transverse geometry of the Nernst effect should enable efficient, large-area and flexible coverage of a heat source. For such applications to be viable, substantial improvements will be necessary not only for their performance but also for the associated material costs, safety and stability. In terms of the electronic structure, the anomalous Nernst effect (ANE) originates from the Berry curvature of the conduction electrons near the Fermi energy<sup>1,2</sup>. To design a large Berry curvature, several approaches have been considered using nodal points and lines in momentum space<sup>3–10</sup>. Here we perform a high-throughput computational search and find that 25 percent doping of aluminium and gallium in alpha iron, a naturally abundant and low-cost element, dramatically enhances the ANE by a factor of more than ten, reaching about 4 and 6 microvolts per kelvin at room temperature, respectively, close to the highest value reported so far. The comparison between experiment and theory indicates that the Fermi energy tuning to the nodal web—a flat band structure made of interconnected nodal lines—is the key for the strong enhancement in the transverse thermoelectric coefficient, reaching a value of about 5 amperes per kelvin per metre with a logarithmic temperature dependence. We have also succeeded in fabricating thin films that exhibit a large ANE at zero field, which could be suitable for designing low-cost, flexible microelectronic thermoelectric generators<sup>11–13</sup>.

Thermoelectricity, the conversion of heat current into electric energy, provides a key technology for versatile energy harvesting and for heat flow sensors. There is a rapidly growing demand for novel energy-harvesting technology to power Internet of Things sensors and wearable devices, particularly in the form of flexible, durable micro-thermoelectric generators ( $\mu$ -TEGs). So far, the technology has relied on the longitudinal thermoelectric response known as the Seebeck effect<sup>11–15</sup>. Recently, its transverse counterpart in ferromagnets, the anomalous Nernst effect (ANE), has gained increasing attention, as it has a number of potential benefits<sup>6–9,16–18</sup>. For example, the transverse geometry of the Nernst effect enables efficient, large-area and flexible coverage of a curved heat source (Fig. 1a, left). It substantially reduces the number of production processes as well as the contact resistance of a thermopile, compared with a conventional thermoelectric device (Fig. 1a, right). In addition, the transverse geometry is hypothetically better suited for thermoelectric conversion as the Ettingshausen heat current should support the Nernst voltage, whereas the Peltier heat current may suppress the Seebeck voltage<sup>18</sup>. However, the ANE is too small compared with the Seebeck effect for any thermoelectric application. Thus, it is important to develop an approach to design a new

class of materials that exhibit a large ANE at zero field. Moreover, as the use of rare and toxic elements would pose a barrier for applications, low-cost and safe elements should be used for thermoelectric materials.

Here we introduce two iron-based cubic compounds  $\text{Fe}_3\text{X}$  (where X is Ga or Al) as materials suitable for designing such low-cost, flexible  $\mu$ -TEGs, in particular by using their thin-film forms. As detailed in Methods, our successful fabrication of their thin films enables us to obtain a large ANE of about  $4 \mu\text{V K}^{-1}$  for  $\text{Fe}_3\text{Ga}$  and about  $2 \mu\text{V K}^{-1}$  for  $\text{Fe}_3\text{Al}$  at room temperature using an in-plane temperature gradient (Extended Data Fig. 3a). In addition, our films of  $\text{Fe}_3\text{Ga}$  and  $\text{Fe}_3\text{Al}$  have in-plane magnetization with a coercivity  $B_c$  of about 40 Oe and 20 Oe, respectively, and exhibit a spontaneous ANE at zero field (Fig. 1c) for the out-of-plane temperature gradient. This enables us to design a much simpler  $\mu$ -TEG than the conventional Seebeck analogues (Fig. 1a, Extended Data Fig. 4)<sup>16</sup>. Moreover, the specific power generation capacity  $\Gamma_p = P_{\text{max}}/(A(\Delta T)^2)$  is estimated to be of the same order as or potentially greater than that of conventional  $\mu$ -TEGs (Methods), where  $P_{\text{max}}$ ,  $A$  and  $\Delta T$  are the maximum power, the area of the thermopile device and the temperature difference, respectively<sup>11–13</sup>. In the following, we discuss

<sup>1</sup>Institute for Solid State Physics, University of Tokyo, Kashiwa, Japan. <sup>2</sup>Department of Physics, University of Tokyo, Tokyo, Japan. <sup>3</sup>CREST, Japan Science and Technology Agency (JST), Honcho Kawaguchi, Japan. <sup>4</sup>Nanomaterials Research Institute, Kanazawa University, Kanazawa, Japan. <sup>5</sup>Center for Emergent Matter Science (CEMS), RIKEN, Wako, Japan. <sup>6</sup>Department of Physics, Tohoku University, Sendai, Japan. <sup>7</sup>Department of Applied Physics, University of Tokyo, Tokyo, Japan. <sup>8</sup>Trans-scale Quantum Science Institute, University of Tokyo, Tokyo, Japan. <sup>9</sup>Institute for Quantum Matter and Department of Physics and Astronomy, Johns Hopkins University, Baltimore, MD, USA. <sup>10</sup>These authors contributed equally: Akito Sakai, Susumu Minami, Takashi Koretsune, Taishi Chen, Tomoya Higo. ✉e-mail: satoru@phys.s.u-tokyo.ac.jp



**Fig. 1 | Transverse thermoelectric conversion, iron compounds and a nodal web.** **a**, Thermopiles based on the ANE (left) and the Seebeck effect (right) have the transverse ( $V \perp \Delta T$ ) and longitudinal ( $V \parallel \Delta T$ ) geometries between the voltage ( $V$ ) and temperature difference ( $\Delta T$ ), respectively. Compared with the pillar structure using p- and n-type semiconductors for the conventional thermoelectrics (right), the thermopile structure using the ANE may be as simple as, for example, a film type (yellow) deposited on a flexible substrate (left)<sup>16</sup>. **b**, Magnetization. **c**, Crystal structures of  $\text{Fe}_3\text{X}$  (where X is Ga or Al) (left) and  $\alpha\text{-Fe}$  (right). **d**, Magnetic field dependence of the Nernst voltage  $-V_{yx}$  for the thin films of  $\text{Fe}_3\text{Ga}$ ,  $\text{Fe}_3\text{Al}$  and  $\alpha\text{-Fe}$ . As  $-V_{yx}$  depends on the sample

geometry and measurement conditions such as  $\Delta T$ , we performed the measurements for all films with the same conditions (thickness, measurement set-up and so on). The temperature gradient in the sample (MgO substrate/ $\text{Fe}_3\text{Ga}$ ,  $\text{Fe}_3\text{Al}$  or  $\alpha\text{-Fe}$ /MgO),  $\nabla T_{\text{all}}$ , is monitored to be  $3.0 \text{ K mm}^{-1}$ . The temperature gradient in the  $\text{Fe}_3\text{Ga}$  ( $\text{Fe}_3\text{Al}$ ) layer is estimated to be  $\nabla T \approx 0.86 \text{ K mm}^{-1}$  ( $1.0 \text{ K mm}^{-1}$ ) (Methods). **d**, Energy dispersion of the conduction band (red) and the valence band (blue) around the L point on the BZ boundary for  $\text{Fe}_3\text{Ga}$ . The nodal web is made of nodal lines, shown by the yellow lines (the same as in Fig. 4d), along which the flat portions of the conduction and the valence bands touch around the L point.

how we discovered these materials, their detailed physical properties using bulk single crystals and the origin of the large ANE.

Recent developments in the understanding of topological aspects in electronic transport led to the discovery of the large anomalous Hall and Nernst effects in spin liquids and antiferromagnetic metals<sup>1,2,6,10,17,19–22</sup>. It also fostered the finding that designing the electronic structure to locate the nodal points called Weyl points near the Fermi energy is crucial to enhance the Berry curvature and thus for ferromagnets to exhibit a large ANE beyond the conventional linear scaling law with magnetization<sup>6–9</sup>. The most popular ferromagnet is elemental alpha iron ( $\alpha\text{-Fe}$ ), which adopts a body-centred cubic (bcc) structure. Although its natural abundance is the highest among the magnetic elements, it does not exhibit a large ANE by itself<sup>17</sup> (Fig. 1c, Extended Data Fig. 3a). It would be useful for applications such as versatile energy harvesting and heat flow sensors if we could find a large ANE in iron-based compounds. To find candidate compounds efficiently, we used a high-throughput computational method to screen materials without synthesizing them first.

The inverse band-structure problem, namely identifying a material with given electronic properties, has a 20-year history<sup>23</sup>. Recently, triggered by the combination of advanced *ab initio* calculation methods and intelligent data-mining techniques for huge databases, *in silico* design has made great progress in a variety of fields in materials science<sup>24–28</sup>. However, high-throughput materials design for transport properties such as the ANE still remains a major challenge, as accurate calculation of not only one-particle quantities (such as the Kohn–Sham energy) but also two-particle quantities (such as the response function) is required. In addition, for the ANE, we need to calculate the Berry curvature using a sufficiently dense  $k$ -mesh in the three-dimensional (3D) Brillouin zone (BZ).

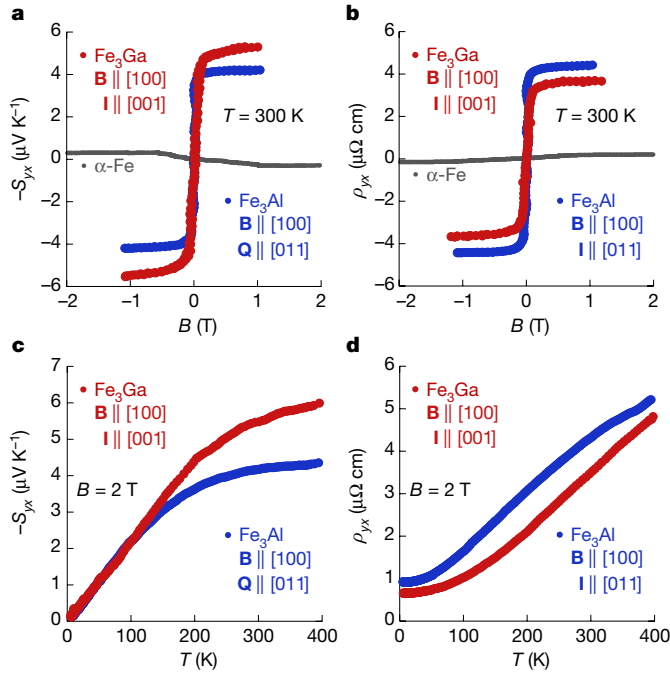
To reduce the formidable numerical cost for materials screening, we recently developed a code (a new version of pw2wannier in Quantum Espresso; Methods) that enables us to perform an efficient Wannier interpolation for calculations with relativistic ultrasoft pseudopotentials. We then performed a calculation for roughly 1,400 magnets available in the Materials Project<sup>28</sup> (Methods, Supplementary Table 1) to provide a comprehensive set of calculations for the transport properties of magnets. Among more than 250 iron-based compounds, we focused on binary compounds from the perspective of future

applications. In Table 1, we list the five most promising candidates that would potentially exhibit a large ANE, with a transverse thermoelectric coefficient  $\alpha_{yx}$  of more than  $2.4 \text{ A K}^{-1} \text{ m}^{-1}$ . Here we select  $\text{Fe}_3\text{Ga}$  as a target material (Fig. 1b, Methods, Extended Data Figs. 1, 2), owing to its large  $|\alpha_{yx}|$  and to its high Curie temperature  $T_C \approx 720 \text{ K}$  (Table 1)<sup>29–33</sup>. In addition, its sister compound  $\text{Fe}_3\text{Al}$  ( $T_C \approx 600 \text{ K}$ ) is also of great interest fundamentally and technologically, as aluminium and iron are the most abundant metal elements in Earth's crust. Furthermore, as both compounds have the same cubic structure (Fig. 1b), we find all the properties to be nearly isotropic, which is convenient for application (Supplementary Information, Extended Data Figs. 6, 8, 9).

First, we present our experimental observation of the giant ANE in  $\text{Fe}_3\text{Ga}$  and  $\text{Fe}_3\text{Al}$  at room temperature. Figure 2a shows that the field dependence of the anomalous Nernst coefficient,  $S_{yx}$ , for  $\text{Fe}_3\text{Ga}$  and  $\text{Fe}_3\text{Al}$  saturates to a value as high as around  $6 \mu\text{V K}^{-1}$  and around  $4 \mu\text{V K}^{-1}$ , respectively, comparable to the value reported for  $\text{Co}_2\text{MnGa}$  (ref. 7). This is striking as it indicates that 25% gallium doping leads to 20-fold enhancement in the ANE, compared with about  $0.3 \mu\text{V K}^{-1}$  for  $\alpha\text{-Fe}$ <sup>17</sup> (Fig. 2a), which shares the same bcc structure as  $\text{Fe}_3\text{Ga}$  and  $\text{Fe}_3\text{Al}$  (Fig. 1b). In addition, the Hall resistivity saturates to about  $3 \mu\Omega \text{ cm}$ , one order magnitude larger than about  $0.2 \mu\Omega \text{ cm}$  for  $\alpha\text{-Fe}$ <sup>17</sup> (Fig. 2b). As shown in Fig. 2c,  $-S_{yx}(T)$  exhibits a  $T$ -linear increase at low temperatures and saturation on heating above room temperature. The Hall resistivity also exhibits a systematic increase on heating for the entire range of temperatures (Fig. 2d).

In general, the anomalous Nernst coefficient is composed of two terms, that is,  $S_{yx} = \alpha_{yx}\rho_{yy} - \sigma_{yx}\rho_{yy}S_{xx}$ , where  $\rho_{yy}$  and  $\sigma_{yx}$  are the longitudinal resistivity and Hall conductivity, respectively (Supplementary Information, Extended Data Figs. 6, 7)<sup>10</sup>. Namely, the transverse charge flow comes from the transverse thermoelectric current  $\alpha_{yx}(-\nabla T)$  and the Hall current. Figure 3a shows the temperature dependence of  $-\alpha_{yx}(T)$ .  $\text{Fe}_3\text{Ga}$  and  $\text{Fe}_3\text{Al}$  share a common feature: a nearly isotropic broad peak following the  $T \ln T$  form, as we will discuss. It is also notable that  $-\alpha_{yx}(T)$  in  $\text{Fe}_3\text{Ga}$  peaks at around 200 K, reaching around  $5.2 \text{ A K}^{-1} \text{ m}^{-1}$ , the largest value for any magnetic compound. By contrast, the Hall conductivity is nearly  $T$  independent for both compounds, ranging between 350 and  $600 \Omega^{-1} \text{ cm}^{-1}$  (Fig. 3b), the size of which is consistent with the scaling law with the longitudinal conductivity  $\sigma_{xx}$  for the intrinsic mechanism of the anomalous Hall effect (Supplementary Information, Extended





**Fig. 2 | Large anomalous Nernst and Hall effects of Fe<sub>3</sub>X.** **a, b**, Magnetic field  $B$  dependence of the ANE  $-S_{yx}$  (**a**) and the anomalous Hall effect  $\rho_{yx}$  (**b**) measured at room temperature. The magnitude of  $B$  in the horizontal axis is corrected for the demagnetization effect. The data for  $\alpha$ -Fe are also plotted for comparison<sup>17</sup>. **c, d**, Temperature dependence of  $-S_{yx}$  (**c**) and  $\rho_{yx}$  (**d**) measured under a magnetic field of 2 T along [100]. **B, Q** and **I** refer to the directions of the magnetic field, heat and electric currents, respectively.

Data Table 1)<sup>1</sup>. As a result,  $S_{yx}$  is dominated by the component  $\alpha_{yx}(-\nabla T)$  for Fe<sub>3</sub>Ga, as shown in Extended Data Fig. 7.

In ferromagnets, the Mott relation  $\alpha_{yx} = -\frac{\pi^2 k_B^2 T}{3|e|} \left( \frac{\partial \sigma_{yx}}{\partial E} \right)_{E=E_F}$  (where  $k_B$ ,  $e$ ,  $E$  and  $E_F$  are the Boltzmann constant, elementary charge, energy and Fermi energy of the electrons, respectively) normally holds at  $T \ll T_C$  and thus  $\alpha_{yx}$  exhibits the Fermi liquid  $T$ -linear dependence. In fact,  $-\alpha_{yx}/T$  shows a nearly  $T$ -constant behaviour at low temperatures for both Fe<sub>3</sub>Ga and Fe<sub>3</sub>Al (Fig. 3a, inset). To further examine the  $T$  dependence, in Fig. 3c we plot  $g(T/T_m) \equiv (\alpha_{yx}/\alpha_{yx}(T_m))(T_m/T)$  versus  $T/T_m$ , where  $g$  is a normalized transverse thermoelectric coefficient and  $T_m$  refers to the temperature where  $|\alpha_{yx}|$  attains its maximum (Extended Data Table 1). Interestingly,  $g$  for both Fe<sub>3</sub>Ga and Fe<sub>3</sub>Al shows a  $\ln T$  dependence up to the highest temperature of our measurements, and shows saturation at large values below  $T/T_m \approx 0.5$ . Similarly, we also plot  $\sigma_{yx}/\sigma_{yx}(T_m)$  for both compounds versus  $T/T_m$  and find that they roughly scale at  $T/T_m > 0.5$  (Fig. 3d).

Using the Berry curvature expression for intrinsic  $\sigma_{yx}$  at  $T=0$  given by  $\sigma_{yx}(E_F) = -(e^2/h) \sum_{n,\mathbf{k}} \Omega_{n,z}(\mathbf{k}) \Theta(E_F - \varepsilon_{n,\mathbf{k}})$ , the Mott relation can be rewritten as  $\alpha_{yx} = \frac{\pi^2 k_B^2 T}{3|e|} \frac{e^2}{h} \sum_{n,\mathbf{k}} \Omega_{n,z}(\mathbf{k}) \delta(E_F - \varepsilon_{n,\mathbf{k}})$ , where the summation is taken for all the occupied states,  $\Omega_{n,z}(\mathbf{k})$  is the Berry curvature along the  $\hat{z}$  direction,  $n$  is the band index,  $\Theta(\varepsilon)$  is the unit step function,  $\delta(\varepsilon)$  is the Dirac delta function and  $h$ ,  $\mathbf{k}$  and  $\varepsilon_{n,\mathbf{k}}$  are Planck's constant, the wavevector and the energy of the electrons in band  $n$ , respectively. Thus, it is essential to find the mechanism to simultaneously enhance the density of states and Berry curvature at the Fermi surfaces to obtain a large  $-\alpha_{yx}/T$  as observed in Fe<sub>3</sub>Ga and Fe<sub>3</sub>Al. Using the Kubo formula, the Berry curvature can be described as  $\Omega_{n,\mu\nu} = i \sum_{n' \neq n} \frac{\langle n | v_\mu | n' \rangle \langle n' | v_\nu | n \rangle}{(\varepsilon_n - \varepsilon_{n'})^2}$ .

Here,  $|n\rangle$  and  $\varepsilon_n$  are the cell-periodic Bloch eigenstates and eigenvalues, and  $v_\mu$  and  $v_\nu$  are velocity operators along directions  $\mu$  and  $\nu$ . The velocity operator is defined as  $v_{\mu(\nu)} = \frac{\partial H(\mathbf{k})}{\partial k_{\mu(\nu)}}$ , using the  $k$ -dependent

**Table 1 | The five most promising iron-based binary compounds obtained from the high-throughput computation**

Formula	Space group	$\alpha_{\max}$ (A K <sup>-1</sup> m <sup>-1</sup> )	$T_c$ (K)
Fe <sub>3</sub> Pt	$Pm\bar{3}m$	6.2	450
Fe <sub>3</sub> Ga	$Fm\bar{3}m$	3.0	720
Fe <sub>3</sub> Al	$Fm\bar{3}m$	2.7	600
Fe <sub>3</sub> Si	$Fm\bar{3}m$	2.5	840
Fe <sub>4</sub> N	$Pm\bar{3}m$	2.4	760

The compounds are sorted by the size of  $\alpha_{\max}$ , that is, the maximum value of  $\alpha_y$ , estimated at the Fermi energy at  $T \leq 500$  K.  $T_c$  is the experimentally obtained Curie temperature<sup>29–33</sup>.

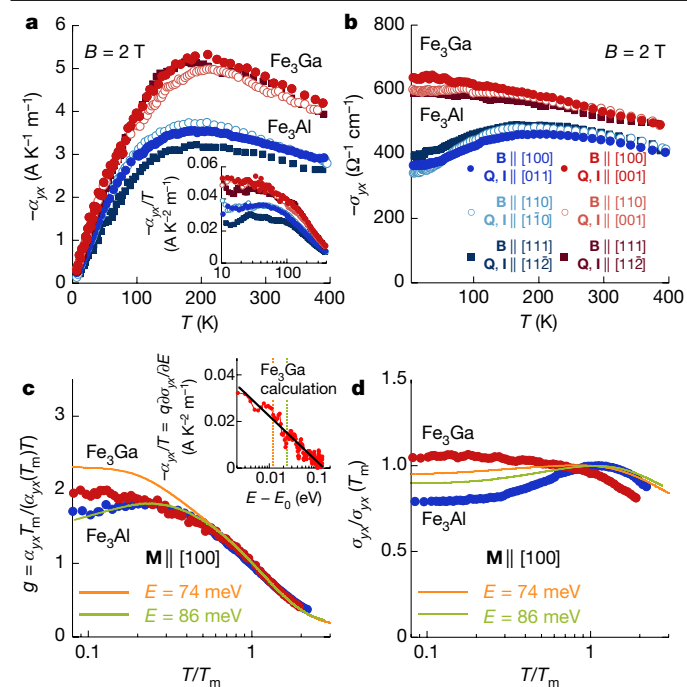
Hamiltonian  $H(\mathbf{k}) = e^{-i\mathbf{k}\cdot\mathbf{r}} H e^{i\mathbf{k}\cdot\mathbf{r}}$  where  $\mathbf{r}$  is the position vector and  $H$  is the original Hamiltonian. This equation indicates that the Berry curvature would be strongly enhanced in the vicinity of the gapped line structure, which originally belongs to the nodal lines before introducing the spin-orbit coupling (SOC) (Fig. 1d). Moreover, the density of states (DOS) would be highly increased when the nodal lines are not only energy independent but also interconnected, forming a quasi-2D network. In the following, we show that such a 'nodal web' structure may arise when two nearly flat valence and conduction bands touch around a highly symmetric point (Figs. 1d, 4b–d) and eventually lead to a large Berry curvature on inclusion of SOC.

Figure 4a shows the band dispersion calculated using density functional theory (DFT) along highly symmetric lines in the momentum space. The numerically obtained  $\sigma_{yx}$  and  $\partial\sigma_{yx}/\partial E$  exhibit a clear kink and an associated peak around 60 meV, respectively (Extended Data Fig. 5a, b). Correspondingly, we find that the temperature dependence of  $\alpha_{yx}/T$  below about 200 K depends strongly on the energy  $E$  around 60 meV. In particular,  $\alpha_{yx}/T$  exhibits the logarithmic temperature dependence over a wide  $T$  range at 62 meV (Extended Data Fig. 5c). Notably, using their own  $T_m$  (Extended Data Table 1), experiment and theory are in agreement when plotting  $g = (\alpha_{yx}/\alpha_{yx}(T_m))(T_m/T)$  versus  $T/T_m$  (Fig. 3c). This indicates that the experimental results can be well reproduced by assuming that  $E$  is located at about  $E_F + 74$  meV for Fe<sub>3</sub>Ga and about  $E_F + 86$  meV for Fe<sub>3</sub>Al. Moreover, the reasonable agreement can be also found in the temperature dependence of the Hall conductivity by using the same set of parameters (Fig. 3d). In addition, theory indicates that the dominant band is made of minority-spin states and thus the upward shifts in  $E$  with respect to  $E_F$  can be naturally understood given that the magnetization is slightly overestimated in theory (Supplementary Information).

Figure 4b shows the 3D network made of the nodal lines calculated for Fe<sub>3</sub>Ga without introducing SOC. All the nodal lines go through the whole BZ and are interconnected with each other, instead of forming a closed loop inside the BZ. In particular, we highlight the nodal-line network with a nearly flat energy dispersion at  $E \approx +74$  meV as the nodal web (red and blue in Fig. 4b, c, yellow lines in Fig. 4d) around the L point. As two mirror eigenstates should be degenerate within each mirror plane (green shaded planes in Fig. 4d), the symmetry-protected six nodal lines branch off from another nodal line connecting the  $\Gamma$  and L points (Fig. 4c).

After introducing SOC, a gap opens along the nodal lines and induces the Berry curvature. We expect that the Berry curvature becomes substantially enhanced near the momenta that originally belong to the flat nodal web, such as the one around the L point. This is because along the nodal line, the gap between the two bands forming the nodal web should be small (Supplementary Information, Extended Data Fig. 5d–f) and, moreover, the flat energy dispersion must enhance the DOS associated with the Berry curvature.

To confirm this, we show the contour plot of the Berry curvature in Fig. 4e. The strongly enhanced Berry curvature is particularly seen at

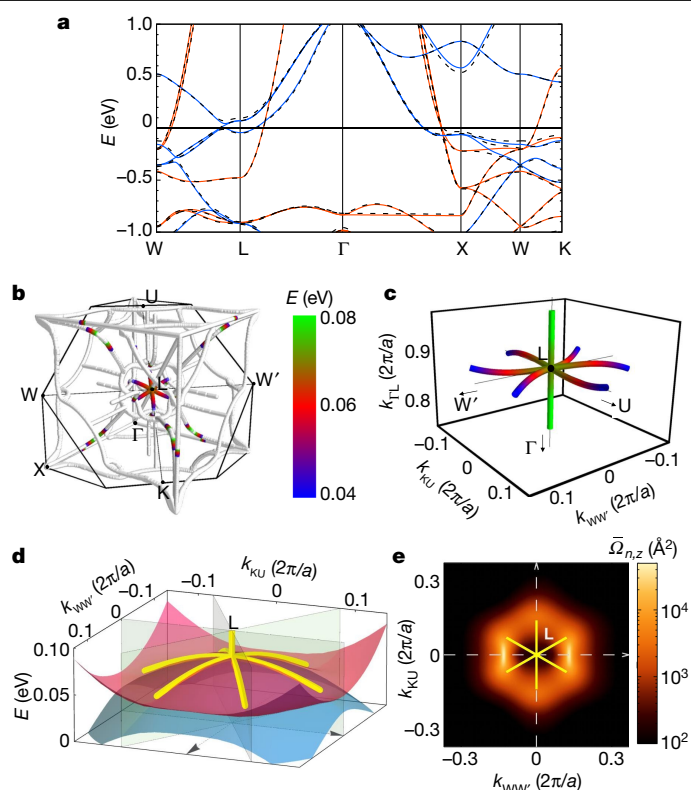


**Fig. 3 | Giant transverse thermoelectric conductivity for  $\text{Fe}_3\text{X}$ .** **a, b**, Temperature dependence of the transverse thermoelectric conductivity  $-\alpha_{yx}$  (**a**) and the Hall conductivity  $-\sigma_{yx}$  (**b**) obtained for various directions of  $\mathbf{B}$ . In **b**, we define the data symbols for the combination of  $\mathbf{Q}$  ( $\mathbf{I}$ ) and  $\mathbf{B}$  directions for **a** and **b**. Both  $-\alpha_{yx}$  and  $-\sigma_{yx}$  exhibit nearly isotropic behaviours. The inset in **a** shows  $-\alpha_{yx}/T$  versus  $T$ . **c, d**, Scaling relations of  $-\alpha_{yx}$  (**c**) and  $\sigma_{yx}$  (**d**) versus  $T/T_m$  for  $\mathbf{M} \parallel [100]$ . Here,  $\mathbf{M}$  and  $T_m$  refer to the magnetization and temperature at which  $|\alpha_{yx}|$  attains its maximum (Extended Data Table 1). The solid lines are obtained from first-principles calculations at  $E = +74$  meV (orange) and  $E = +86$  meV (green). The inset in **c** shows the  $-\ln|E - E_0|$  dependence (black solid line) of  $\frac{\alpha_{yx}}{T} = -q \frac{\partial \sigma_{yx}}{\partial E}$  with  $q = \frac{\pi^2 k_B^2}{3|e|}$  calculated in the vicinity of  $E_0 = +62$  meV above the Fermi energy. The vertical broken lines indicate the characteristic energies  $E = 74$  meV (orange) and  $E = 86$  meV (green). Nearly isotropic character is found for  $\mathbf{M} \parallel [100]$ ,  $[110]$  and  $[111]$  (Supplementary Information, Extended Data Fig. 9).

the momenta connecting the edge of the nodal web around the L point on the BZ boundary, extending over a quasi-2D area spanned by the web. This situation could be comparable to the case of the band edge of the 1D system being accompanied by a logarithmic increase of the DOS as a function of energy. Similarly,  $\partial \sigma_{yx} / \partial E$  exhibits a logarithmic energy dependence near  $E = 62$  meV, indicating a divergent behaviour of the Berry curvature (Fig. 3c, inset). This leads to the  $-\ln T$  dependence in  $\alpha_{yx}/T$  found at 62 meV (Fig. 3c, Extended Data Fig. 5), the breakdown of the Mott relation. Such  $-\ln T$  dependence has been previously discussed for the Lifshitz quantum critical point between type I and type II Weyl fermions found for  $\text{Co}_2\text{MnGa}$  (ref. 7). Our scenario indicates that even without having Weyl points nearby, the formation of a flat nodal web may lead to the logarithmic enhancement of the transverse thermoelectric conductivity  $\alpha_{yx}$ .

Our discovery of the large spontaneous transverse thermoelectric effects indicates that two iron-based compounds,  $\text{Fe}_3\text{Ga}$  and  $\text{Fe}_3\text{Al}$ , should be suitable for designing low-cost, flexible  $\mu$ -TEGs by using their thin-film form. To gain a stable performance in daily use, it is important to enhance the coercivity further. Finally, it would be also an interesting future direction to look for an enhanced output by combining the ANE with the spin Seebeck effect, both of which occur in the same transverse geometry<sup>34–37</sup>.

**Note added in proof:** During the review process of this paper, we became aware of a work by Nakayama et al.<sup>38</sup>, the received date of which is one week later than ours. While they observed an enhancement in



**Fig. 4 | Evidence for the nodal web structure.** **a**, Band structure around the Fermi energy. Red (blue) lines correspond to up (down) spin bands computed without SOC. Dashed lines show the bands calculated including SOC. **b**, BZ showing the nodal-line network. The nodal web formed around the L point is highlighted by the colour, which shows the energy dispersion of the nodal lines on a linear scale in a narrow range between 0.04 eV (blue) and 0.08 eV (green) above the Fermi energy. **c**, Three-dimensional network forming the nodal web around the L point. The  $k_{\text{WW}}$ ,  $k_{\text{KU}}$  and  $k_{\text{FL}}$  axes indicate the directions parallel to  $\text{W}-\text{W}'$ ,  $\text{K}-\text{U}$  and  $\Gamma-\text{L}$ , respectively. The nodal line (green) connecting the  $\Gamma$  and L points branches off into three nodal lines (red-blue) at the L point. **d**, Nearly flat energy dispersion of the nodal lines (yellow) forming the web structure around the L point, on the plane perpendicular to  $\Gamma-\text{L}$ . The  $k_{\text{WW}}$  and  $k_{\text{KU}}$  axes are the same as those in **c** and the  $z$  axis indicates the energy. A higher-energy (red) and a lower-energy (light blue) band touches with each other sharing gapless nodal lines (yellow lines, the same as in Fig. 1c) along the mirror planes (shaded green planes) in the case without SOC. **e**, Contour plot of the Berry curvature  $\bar{\Omega}_{n,z}$  of the lower-energy band  $n$  (in **d**) around the L point and nodal lines. The  $z$  direction is along  $[001]$  parallel to the magnetization. The Berry curvature was integrated within  $[-\Delta k, \Delta k]$  along the  $k_{\text{FL}}$  axis;  $\bar{\Omega}_{n,z}(k_{\text{WW}}, k_{\text{KU}}) = \int_{-\Delta k}^{\Delta k} dk_{\text{FL}} \Omega_{n,z}(k_{\text{WW}}, k_{\text{KU}}, k_{\text{FL}})$ .  $\Delta k$  was chosen to be sufficient to include the nodal web structure. The Berry curvature is strongly enhanced in the extended region around the nodal web structure (yellow lines) (Methods, Extended Data Fig. 5d–f).

the ANE by gallium doping in bcc-Fe in its thin-film form, the materials used in their study are in a disordered phase, not the ordered  $\text{DO}_3$  phase as studied in this work. Moreover, the maximum of  $|\sigma_{yx}|$  in Nakayama et al.<sup>38</sup> is less than half of the values found in our experiment and does not appear at  $x = 0.25$  (that is,  $\text{Fe}_3\text{Ga}$ ). As a result, the transverse thermoelectric conductivity  $\alpha_{yx}$  is about five times smaller than our estimate, highlighting the importance of both the ordered  $\text{DO}_3$  crystal structure and the stoichiometry of  $\text{Fe}_3\text{Ga}$ . Our work also contains the experimental results using not only thin films but also bulk single crystals over a wide temperature range with different field orientations. The combination of our comprehensive experimental observations and theoretical analysis have revealed that a ‘nodal web’ structure in the vicinity of the Fermi energy plays an important role in enhancing

the Berry curvature and thereby the transverse thermoelectric effect. In addition to the in-plane heat current configuration, we performed ANE measurements with the heat current applied perpendicular to the thin film, which reveals the zero-field spontaneous anomalous Nernst voltage with a coercivity of about 40 Oe in Fe<sub>3</sub>Ga.

## Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-020-2230-z>.

- Nagaosa, N., Sinova, J., Onoda, S., MacDonald, A. H. & Ong, N. P. Anomalous Hall effect. *Rev. Mod. Phys.* **82**, 1539–1592 (2010).
- Xiao, D., Chang, M.-C. & Niu, Q. Berry phase effects on electronic properties. *Rev. Mod. Phys.* **82**, 1959–2007 (2010).
- Wan, X., Turner, A. M., Vishwanath, A. & Savrasov, S. Y. Topological semimetal and Fermi arc surface states in the electronic structure of pyrochlore iridates. *Phys. Rev. B* **83**, 205101 (2011).
- Burkov, A. A. & Balents, L. Weyl semimetal in a topological insulator multilayer. *Phys. Rev. Lett.* **107**, 127205 (2011).
- Fang, C., Chen, Y., Kee, H.-Y. & Fu, L. Topological nodal line semimetals with and without spin-orbital coupling. *Phys. Rev. B* **92**, 081201 (2015).
- Ikhtlas, M. et al. Large anomalous Nernst effect at room temperature in a chiral antiferromagnet. *Nat. Phys.* **13**, 1085–1090 (2017).
- Sakai, A. et al. Giant anomalous Nernst effect and quantum-critical scaling in a ferromagnetic semimetal. *Nat. Phys.* **14**, 1119–1124 (2018).
- Noky, J., Xu, Q., Felser, C. & Sun, Y. Large anomalous Hall and Nernst effects from nodal line symmetry breaking in Fe<sub>2</sub>MnX (X = P, As, Sb). *Phys. Rev. B* **99**, 165117 (2019).
- Guin, S. N. et al. Zero-field Nernst effect in a ferromagnetic kagome-lattice Weyl-semimetal Co<sub>5</sub>Sn<sub>2</sub>S<sub>2</sub>. *Adv. Mater.* **31**, 1806622 (2019).
- Lee, W.-L., Watauchi, S., Miller, V. L., Cava, R. J. & Ong, N. P. Anomalous Hall heat current and Nernst effect in the CuCr<sub>2</sub>Se<sub>4-x</sub>Br<sub>x</sub> ferromagnet. *Phys. Rev. Lett.* **93**, 226601 (2004).
- Bottner, H., Nurnus, J., Schubert, A. & Volkert, F. New high density micro structured thermogenerators for stand alone sensor systems. In *Proc. 26th Int. Conf. Thermoelectrics* 306–309 (IEEE, 2007).
- Glatz, W., Schwyter, E., Durrer, L. & Hierold, C. Bi<sub>2</sub>Te<sub>3</sub>-based flexible micro thermoelectric generator with optimized design. *J. Microelectromech. Syst.* **18**, 763–772 (2009).
- Hu, G., Edwards, H. & Lee, M. Silicon integrated circuit thermoelectric generators with a high specific power generation capacity. *Nat. Electron.* **2**, 300–306 (2019).
- Bell, L. E. Cooling, heating, generating power, and recovering waste heat with thermoelectric systems. *Science* **321**, 1457–1461 (2008).
- Snyder, G. J. & Toberer, E. S. Complex thermoelectric materials. *Nat. Mater.* **7**, 105–114 (2008).
- Sakuraba, Y. et al. Anomalous Nernst effect in L<sub>10</sub>-FePt/MnGa thermopiles for new thermoelectric applications. *Appl. Phys. Express* **6**, 033003 (2013).
- Li, X. et al. Anomalous Nernst and Righi-Leduc effects in Mn<sub>3</sub>Sn: Berry curvature and entropy flow. *Phys. Rev. Lett.* **119**, 056601 (2017).
- Mizuguchi, M. & Nakatsuji, S. Energy-harvesting materials based on the anomalous Nernst effect. *Sci. Technol. Adv. Mater.* **20**, 262–275 (2019).
- Machida, Y., Nakatsuji, S., Onoda, S., Tayama, T. & Sakakibara, T. Time-reversal symmetry breaking and spontaneous Hall effect without magnetic dipole order. *Nature* **463**, 210–213 (2010).
- Nakatsuji, S., Kiyohara, N. & Higo, T. Large anomalous Hall effect in a noncollinear antiferromagnet at room temperature. *Nature* **527**, 212–215 (2015).
- Kiyohara, N., Tomita, T. & Nakatsuji, S. Giant anomalous Hall effect in the chiral antiferromagnet Mn<sub>3</sub>Ge. *Phys. Rev. Appl.* **5**, 064009 (2016).
- Nayak, A. K. et al. Large anomalous Hall effect driven by a nonvanishing Berry curvature in the noncollinear antiferromagnet Mn<sub>3</sub>Ge. *Sci. Adv.* **2**, e1501870 (2016).
- Franceschetti, A. & Zunger, A. The inverse band-structure problem of finding an atomic configuration with given electronic properties. *Nature* **402**, 60–63 (1999).
- Mounet, N. et al. Two-dimensional materials from high-throughput computational exfoliation of experimentally known compounds. *Nat. Nanotechnol.* **13**, 246–252 (2018).
- Curtarolo, S. et al. The high-throughput highway to computational materials design. *Nat. Mater.* **12**, 191–201 (2013).
- Saal, J. E., Kirklin, S., Aykol, M., Meredig, B. & Wolverton, C. Materials design and discovery with high-throughput density functional theory: the open quantum materials database (OQMD). *JOM* **65**, 1501–1509 (2013).
- Curtarolo, S. et al. AFLOW: an automatic framework for high-throughput materials discovery. *Comput. Mater. Sci.* **58**, 218–226 (2012).
- Jain, A. et al. A high-throughput infrastructure for density functional theory calculations. *Comput. Mater. Sci.* **50**, 2295–2310 (2011).
- Sumiyama, K., Emoto, Y., Shiga, M. & Nakamura, Y. Magnetic and magnetovolume properties of the Cu<sub>2</sub>Au type ordered Fe-Pt alloys around the γ-α phase boundary. *J. Phys. Soc. Jpn* **50**, 3296–3302 (1981).
- Kawamiya, N., Adachi, K. & Nakamura, Y. Magnetic properties and Mössbauer investigations of Fe-Ga alloys. *J. Phys. Soc. Jpn* **33**, 1318–1327 (1972).
- Shinohara, T. The effect of atomic ordering on the magnetic properties of Fe-Al alloys. *J. Phys. Soc. Jpn* **19**, 51–58 (1964).
- Niculescu, V. et al. Relating structural, magnetization, and hyperfine field studies to a local environment model in Fe<sub>3-x</sub>V<sub>x</sub>Si and Fe<sub>3-x</sub>Mn<sub>x</sub>Si. *Phys. Rev. B* **14**, 4160–4176 (1976).
- Guilland, C. & Creveaux, H. Preparation and magnetic properties of the compound Fe<sub>4</sub>N. *Compt. Rend. Acad. Sci.* **222**, 1170–1173 (1946).
- Uchida, K. et al. Observation of the spin Seebeck effect. *Nature* **455**, 778–781 (2008).
- Ramos, R. et al. Observation of the spin Seebeck effect in epitaxial Fe<sub>3</sub>O<sub>4</sub> thin films. *Appl. Phys. Lett.* **102**, 072413 (2013).
- Boona, S. R., Vandaele, K., Boona, I. N., McComb, D. W. & Heremans, J. P. Observation of spin Seebeck contribution to the transverse thermopower in Ni-Pt and MnBi-Au bulk nanocomposites. *Nat. Commun.* **7**, 13714 (2016).
- Kannan, H., Fan, X., Celik, H., Han, X. & Xiao, J. Q. Thickness dependence of anomalous Nernst coefficient and longitudinal spin Seebeck effect in ferromagnetic Ni<sub>x</sub>Fe<sub>100-x</sub> films. *Sci. Rep.* **7**, 6175 (2017).
- Nakayama, H. et al. Mechanism of strong enhancement of anomalous Nernst effect in Fe by Ga substitution. *Phys. Rev. Mater.* **3**, 114412 (2019).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2020



### Single-crystal growth and experimental methods

Single crystals of  $\text{Fe}_3\text{X}$  (where X is Ga or Al) were prepared by the Czochralski method after making polycrystalline samples by arc melting. As-grown single crystals were used for all the measurements. Our analyses using both inductively coupled plasma spectroscopy and energy dispersive X-ray analysis indicate that our single crystals are stoichiometric within a few percent resolution. The samples were oriented by the Laue backscattering method and subsequently cut by spark erosion. The bar-shaped sample with flat polished surfaces was used for all the transport measurements, including the electric resistivity and Hall, Seebeck and Nernst effects in a commercial system (thermal transport option (TTO) of physical property measurement system (PPMS), Quantum Design). A schematic of the measurement setup is provided in Extended Data Fig. 3b. The three gold-plated copper leads were attached to the sides of a rectangular parallelepiped sample with silver epoxy, which were connected to one resistive heater and two (hot and cold) Cernox1050 thermometers by screws. Two copper wires were also attached on the side surfaces of the sample with silver epoxy, which were used to measure the Hall and Nernst voltages. One of the edges of the sample was directly clamped to the heat sink of the TTO puck by the screw. The magnetic field  $\mathbf{B}$  was applied along [100], [110] or [111]. The electric (thermal) current  $\mathbf{I}$  ( $\mathbf{Q}$ ) was applied perpendicular to  $\mathbf{B}$ , namely  $\mathbf{I}$  ( $\mathbf{Q}$ )  $\parallel$  [001] or [011] for  $\mathbf{B} \parallel$  [100], [001] or  $\bar{1}\bar{1}0$  for  $\mathbf{B} \parallel$  [110], and [112] for  $\mathbf{I}$  ( $\mathbf{Q}$ )  $\parallel$  [111] without a transverse component (adiabatic Nernst effect). All the transport properties represent steady-state values, which were taken under a constant heat/electric current. Magnetization was measured by using a commercial SQUID magnetometer (MPMS, Quantum Design) after reshaping the single crystal into a tiny piece (about 0.1 mg). During the measurement, the magnetic field was fixed at each measurement point in the persistent mode.

The anisotropy in the Nernst and Hall effects and magnetization was found to be small, as we discuss in Supplementary Information (Extended Data Figs. 6, 8, 9).

### Crystal structure

$\text{Fe}_3\text{X}$  has a bcc derivative structure, namely, an ordered  $D0_3$  structure ( $Fm\bar{3}m$ ), partially ordered  $B_2$  structure ( $Pm\bar{3}m$ ), which corresponds to stoichiometric  $\text{Fe}_3\text{X}$  phase, and fully disordered  $A_2$  structure ( $Im\bar{3}m$ ). Our single crystals were confirmed to have a  $D0_3$  structure by X-ray diffraction (XRD) (Extended Data Fig. 1a, b) and selected area electron diffraction (Extended Data Fig. 1c). The XRD pattern is well reproduced by the Rietveld analysis as shown by the solid line in Extended Data Fig. 1a, b, indicating the single phase of the  $D0_3$   $\text{Fe}_3\text{X}$  with a lattice constant of  $a = 5.80 \text{ \AA}$  for  $\text{Fe}_3\text{Ga}$  and  $a = 5.79 \text{ \AA}$  for  $\text{Fe}_3\text{Al}$ . For  $\text{Fe}_3\text{Al}$ , (111) peak is clearly observed around  $2\theta \approx 27^\circ$  (where  $2\theta$  is the angle between incident and diffracted X-ray beam directions), indicating the  $D0_3$  ordered structure. For  $\text{Fe}_3\text{Ga}$ , the (111) peak is expected to be very small (about 0.8% of the (220) peak) and cannot be observed in our experiment. Nevertheless, we detect the signal from (111) by the electron diffraction as shown in Extended Data Fig. 1c. While the (111) and (200) reflections are detected by a charge-coupled device camera for  $\text{Fe}_3\text{Al}$ , they are detected by only an imaging plate with long exposure time for  $\text{Fe}_3\text{Ga}$ . Note that the effect of multiple reflection and diffraction from the oxide appears due to the high sensitivity of the imaging plate and long exposure time. For the powder XRD measurements, we annealed the sample at  $630^\circ\text{C}$  for about 10 h for  $\text{Fe}_3\text{Ga}$ , and  $400^\circ\text{C}$  for about 20 h for  $\text{Fe}_3\text{Al}$  after making powder from single crystals to remove the strains and distortions introduced by the crushing procedure.

### Fabrication and characterization for $\text{Fe}_3\text{Al}$ and $\text{Fe}_3\text{Ga}$ thin films

$\text{Fe}_3\text{Al}$  films were fabricated on  $\text{MgO}(001)$  substrates by a molecular beam epitaxy method under ultrahigh vacuum (UHV). The base pressure of the UHV chamber was below  $2 \times 10^{-8}$  Pa. Before the deposition,

the substrates were annealed at  $800^\circ\text{C}$  for 10 min in the UHV chamber. An  $\text{MgO}$  layer (5 nm) was deposited at a rate of  $0.1 \text{ \AA s}^{-1}$  at room temperature on the  $\text{MgO}$  substrate as a seed layer for the  $\text{Fe}_3\text{Al}$  layer. The  $\text{Fe}_3\text{Al}$  layer (50 nm) was deposited at a rate of  $0.2 \text{ \AA s}^{-1}$  at room temperature and subsequently it was annealed at  $450^\circ\text{C}$  for 30 min.

$\text{Fe}_3\text{Ga}$  films were fabricated on  $\text{MgO}(001)$  substrates by a d.c. magnetron sputtering method. The base pressure of the chamber was below  $5 \times 10^{-7}$  Pa. Before the deposition, the substrates were annealed at  $800^\circ\text{C}$  for 10 min in the chamber. An  $\text{Fe}_3\text{Ga}$  layer (50 nm) was deposited at a rate of  $0.4 \text{ \AA s}^{-1}$  at room temperature and subsequently it was annealed at  $500^\circ\text{C}$  for 30 min.

The compositions of the  $\text{Fe}_3\text{Al}$  and  $\text{Fe}_3\text{Ga}$  layers were estimated to be  $\text{Fe}_{2.7}\text{Al}_{1.3}$  and  $\text{Fe}_{3.07}\text{Ga}_{0.93}$ , respectively, by scanning electron microscopy-energy dispersive X-ray spectroscopy. To prevent oxidation, the  $\text{Fe}_3\text{Al}$  and  $\text{Fe}_3\text{Ga}$  layers were both capped with  $\text{MgO}$  (5 nm) under the same conditions as described for the seed layer.

The structural properties of the  $\text{Fe}_3\text{Al}$  and  $\text{Fe}_3\text{Ga}$  films were analysed by XRD measurements using a  $\text{Cu-K}\alpha$  source. Extended Data Fig. 2a shows the XRD patterns ( $2\theta/\omega$ -scan) of the  $\text{Fe}_3\text{Al}$  and  $\text{Fe}_3\text{Ga}$  films made by the solid phase epitaxy. We confirmed the (002) and (004) peaks in the  $\text{Fe}_3\text{Al}$  and  $\text{Fe}_3\text{Ga}$  films indexed by the Bravais lattice of the bcc structure with the lattice constants of  $c = 5.71 \text{ \AA}$  and  $c = 5.74 \text{ \AA}$ , respectively. Extended Data Fig. 2b presents the XRD patterns ( $\varphi$ -scan) of the {202} planes of the  $\text{Fe}_3\text{Al}$  and  $\text{Fe}_3\text{Ga}$  films, and the  $\text{MgO}$  substrate. The reflections of both  $\text{Fe}_3\text{Al}$  and  $\text{Fe}_3\text{Ga}$  films show fourfold symmetry with  $90^\circ$  intervals. These XRD patterns indicate that these layers are grown epitaxially with a well-defined in-plane orientation. As the peaks of  $\text{Fe}_3\text{X}$  {202} are shifted by  $45^\circ$  with respect to the corresponding {202} peaks of  $\text{MgO}$ , the crystallographic relationship is determined to be  $\text{Fe}_3\text{X}(001)[110] \parallel \text{MgO}(001)[100]$ . The presence of the superlattice (111) peak was further examined to determine whether the films are chemically ordered in the  $D0_3$  type. Extended Data Fig. 2c shows the (111) peaks in the XRD pattern ( $2\theta/\omega$ -scan) for the  $\text{Fe}_3\text{Al}$  and  $\text{Fe}_3\text{Ga}$  thin films, indicating the  $D0_3$  structure.

### Magneto-thermoelectric transport properties of $\text{Fe}_3\text{Al}$ and $\text{Fe}_3\text{Ga}$ thin films

To estimate the anomalous Nernst coefficient  $S_{yx}$  of the  $\text{Fe}_3\text{Ga}$  and  $\text{Fe}_3\text{Al}$  thin films, we performed the Nernst voltage measurements with the temperature gradient  $\nabla T$  along both the in-plane and the out-of-plane directions. For the measurements using the former ‘in-plane’  $\nabla T$  configuration, we employed the same apparatus (TTO of PPMS) and the same temperature gradient configuration as used to measure the bulk  $\text{Fe}_3\text{Ga}$  and  $\text{Fe}_3\text{Al}$  samples (see ‘Single-crystal growth and experimental methods’ and Extended Data Fig. 3b). Note that  $\nabla T$  in the thin film is determined by the  $\nabla T$  in the substrate for the ‘in-plane’  $\nabla T$  configuration, as the thermal resistance of the sample ( $\text{MgO}$  substrate (500  $\mu\text{m}$ )/ $\text{Fe}_3\text{X}$  (50 nm)/ $\text{MgO}$  capping layer (5 nm)) is dominated by the  $\text{MgO}$  substrate as follows. The total thermal resistance  $\theta_{\text{total}}$  is described as  $1/\theta_{\text{total}} = 1/\theta_{\text{substrate}} + 1/\theta_{\text{Fe}_3\text{X}} + 1/\theta_{\text{cap}}$ , where  $\theta_{\text{substrate}}$ ,  $\theta_{\text{Fe}_3\text{X}}$  and  $\theta_{\text{cap}}$  are the thermal resistance of the  $\text{MgO}$  substrate including the seed layer,  $\text{Fe}_3\text{X}$  (50 nm) and  $\text{MgO}$  capping layer, respectively. Each thermal resistance can be estimated by using the sample dimensions as described in Extended Data Fig. 3b as well as the thermal conductivity of  $\text{MgO}$  and  $\text{Fe}_3\text{X}$ ,  $\kappa_{\text{MgO}} \approx 60 \text{ W K}^{-1} \text{ m}^{-1}$  and  $\kappa_{\text{Fe}_3\text{X}} \approx 20 \text{ W K}^{-1} \text{ m}^{-1}$  (refs. <sup>39,40</sup>). As  $\theta_{\text{substrate}} \approx 150 \text{ K W}^{-1}$  is much smaller than both  $\theta_{\text{Fe}_3\text{X}} \approx 4.5 \times 10^6 \text{ K W}^{-1}$  and  $\theta_{\text{cap}} \approx 1.5 \times 10^7 \text{ K W}^{-1}$ ,  $1/\theta_{\text{Fe}_3\text{X}}$  and  $1/\theta_{\text{cap}}$  are negligibly small and thus we do not need to consider any contributions other than the one from the  $\text{MgO}$  substrate to estimate the in-plane temperature gradient of the sample. Independently, we carried out the thermal conductivity measurement of the same sample and confirmed that the thermal conductivity is estimated to be  $56 \text{ W K}^{-1} \text{ m}^{-1}$ , which is consistent with the above estimate based on literature<sup>39</sup>.

Extended Data Fig. 3a shows the field dependence of the anomalous Nernst coefficient for the in-plane temperature gradient  $\nabla T$  (heat

current  $\mathbf{Q} \parallel [1\bar{1}0]$ ) and the perpendicular magnetic field ( $\mathbf{B} \parallel [001]$ ) at 300 K. The Nernst coefficient is estimated to be  $|S_{yx}| = 4.0 \mu\text{V K}^{-1}$ ,  $1.8 \mu\text{V K}^{-1}$  and  $0.3 \mu\text{V K}^{-1}$  for the  $\text{Fe}_3\text{Ga}$ ,  $\text{Fe}_3\text{Al}$  and  $\alpha\text{-Fe}$  (50 nm) thin films, respectively, consistent with the bulk results. Note that the  $\alpha\text{-Fe}$  (50 nm) film was epitaxially grown by the molecular beam epitaxy method on the  $\text{MgO}(001)$  substrate with the same in-plane orientation as the  $\text{Fe}_3\text{X}$  films, and the linear-in- $B$  ordinary Nernst effect  $N \approx -0.05 \mu\text{V K}^{-1}$  is subtracted for the  $\alpha\text{-Fe}$  sample.

In addition to the abovementioned 'in-plane'  $\nabla T$  and out-of-plane  $\mathbf{B}$  configuration, we performed the Nernst effect measurements with 'out-of-plane'  $\nabla T$  and in-plane  $\mathbf{B}$ , which is a more practical configuration for the magneto-thermoelectric devices.

The films used for the 'out-of-plane'  $\nabla T$  measurements are sandwiched from top to bottom, in the following order, as illustrated in Extended Data Fig. 3c: [(1) Resistive heater] + [(2) Copper heat sink] + [(3) Thermal interface silicone soft pad (TIP) (500  $\mu\text{m}$ )] + [(4)  $\text{MgO}$  capping layer (5 nm)] + [(5) Sample:  $\text{Fe}_3\text{Ga}$ ,  $\text{Fe}_3\text{Al}$  or  $\alpha\text{-Fe}$  (50 nm)] + [(6)  $\text{MgO}$  substrate (500  $\mu\text{m}$ )] + [(7) TIP (500  $\mu\text{m}$ )] + [(8) Copper heat sink]. Here we used a thermal interface silicone pad TC-50HSV-1.4 (thermal conductivity,  $\kappa \approx 1.3 \text{ W K}^{-1} \text{ m}^{-1}$ ; thermal resistance, about  $1.1 \text{ K W}^{-1}$ ; Shin-Etsu Silicone) for both TIP layers. In the measurements, the magnetic field and heat current were respectively applied to  $[110]$  and  $[001]$  directions of each film. As the stacking structure is complicated, it is important to make further confirmations of the Nernst effects using several independent methods to estimate the temperature gradient/difference.

Figure 1c shows the field dependence of the Nernst voltage of the  $\text{Fe}_3\text{X}$  and  $\alpha\text{-Fe}$  thin films at room temperature.  $\Delta T_{\text{all}}$  is monitored to be 1.5 K in all the samples by two calibrated thermocouples placed on the top and bottom surfaces of the samples (Extended Data Fig. 3c). The  $\text{Fe}_3\text{Ga}$  ( $\text{Fe}_3\text{Al}$ ) sample shows a sharp hysteresis with a zero-field value of 19.8  $\mu\text{V}$  (10.6  $\mu\text{V}$ ) and coercivity of 40 Oe (20 Oe). These signals are much larger than those obtained for  $\alpha\text{-Fe}$ . An isotropic character of the anomalous Nernst coefficient, as confirmed in the bulk  $\text{Fe}_3\text{X}$  measurements (Supplementary Information, Extended Data Fig. 8), allows us to adopt the in-plane Nernst coefficient (Extended Data Fig. 3a) to estimate the applied out-of-plane temperature difference. The temperature difference and temperature gradient of the  $\text{Fe}_3\text{Ga}$  ( $\text{Fe}_3\text{Al}$ ) layer are estimated to be about  $4.5 \times 10^{-5} \text{ K}$  and  $0.9 \text{ K mm}^{-1}$  ( $5.4 \times 10^{-5} \text{ K}$  and  $1.1 \text{ K mm}^{-1}$ ), respectively. These values are consistent with the temperature difference and gradient estimated by another independent method, as we will discuss below.

To estimate the temperature difference and gradient applied in the  $\text{Fe}_3\text{X}$  film (50 nm), the Nernst effect of an  $\alpha\text{-Fe}$  epitaxial thin film (50 nm) was measured under the same conditions as in the case of the  $\text{Fe}_3\text{Al}$  film. By assuming that the  $\alpha\text{-Fe}$  film has the same anomalous Nernst coefficient  $S_{yx}$  and thermal conductivity as those reported for the bulk  $\alpha\text{-Fe}$  ( $S_{yx,\alpha\text{-Fe}} = 0.3 \mu\text{V K}^{-1}$  (ref. 17) and  $\kappa_{\alpha\text{-Fe}} = 80 \text{ W K}^{-1} \text{ m}^{-1}$  (ref. 41)), the temperature gradient, temperature difference and heat flux of the  $\alpha\text{-Fe}$  film (50 nm) can be estimated to be  $\nabla T_{\alpha\text{-Fe}} = 0.19 \text{ K mm}^{-1}$ ,  $\Delta T_{\alpha\text{-Fe}} = 1.0 \times 10^{-5} \text{ K}$  and  $Q_{\alpha\text{-Fe}} \approx 0.21 \text{ W}$ , respectively.

As we use the same stacking sequence and temperature gradient (difference) for both  $\alpha\text{-Fe}$  and  $\text{Fe}_3\text{X}$  on  $\text{MgO}$  (that is,  $\nabla T_{\text{all}} = 3.0 \text{ K mm}^{-1}$ ,  $\Delta T_{\text{all}} = 1.5 \text{ K}$ ), the heat current of  $\text{Fe}_3\text{X}$  is expected to be almost the same, that is,  $Q_{\text{Fe}_3\text{X}} \approx 0.21 \text{ W}$ . By using the thermal conductivity  $\kappa_{\text{Fe}_3\text{Ga}}$  ( $\kappa_{\text{Fe}_3\text{Al}} = 19.0$  (17.4)  $\text{W K}^{-1} \text{ m}^{-1}$  of  $\text{Fe}_3\text{X}$  presented in Extended Data Fig. 6 (Supplementary Information), the temperature gradient and temperature difference applied in the  $\text{Fe}_3\text{Ga}$  ( $\text{Fe}_3\text{Al}$ ) film can be estimated to be  $\nabla T_{\text{Fe}_3\text{Ga}} = 0.82 \text{ K mm}^{-1}$  and  $\Delta T_{\text{Fe}_3\text{Ga}} = 4.1 \times 10^{-5} \text{ K}$  ( $\nabla T_{\text{Fe}_3\text{Al}} = 0.89 \text{ K mm}^{-1}$  and  $\Delta T_{\text{Fe}_3\text{Al}} = 4.5 \times 10^{-5} \text{ K}$ ), respectively. These are nearly the same values as those calculated using the Nernst voltage (Fig. 1c) and the Nernst coefficients obtained in the 'in-plane'  $\nabla T$  configuration (Extended Data Fig. 3a), indicating that our method to estimate the temperature gradient is reasonable. Another way of confirming the validity for the 'out-of-plane'  $\nabla T$  measurement is to examine the Nernst coefficients using the above estimates of the temperature difference and the Nernst

voltage. The obtained Nernst coefficients are estimated to be  $S_{yx} = 4.4 \mu\text{V K}^{-1}$  and  $2.2 \mu\text{V K}^{-1}$  for the  $\text{Fe}_3\text{Ga}$  and  $\text{Fe}_3\text{Al}$  50-nm thin films, respectively, which are nearly the same values as those found using the Nernst effect measurements in the 'in-plane'  $\nabla T$  configuration (Extended Data Fig. 3a).

### Estimation of the specific power generation capacity

$$\Gamma_p = P_{\text{max}}/(A(\Delta T)^2)$$

In this section, we evaluate how much power can be obtained from the  $\mu\text{-TEG}$  using the anomalous Nernst effect of  $\text{Fe}_3\text{X}$ . Extended Data Fig. 4 shows a schematic of the  $\mu\text{-TEG}$  based on the anomalous Nernst effect<sup>16</sup>. The thermopile consists of an alternating array of the  $\text{Fe}_3\text{X}$  (yellow) and electrode (gold) wires placed on a substrate. Two-hundred wires, whose thickness, width and length are designed to be  $t = 1 \mu\text{m}$ ,  $w = 20 \mu\text{m}$  and  $l = 9 \text{ mm}$ , respectively, are aligned in the substrate with the area of  $A = 1 \text{ cm}^2$ . Owing to the lateral configuration of the anomalous Nernst effect, a temperature difference  $\Delta T$  is applied perpendicular to the plane and the resultant voltage is  $V/\Delta T = LS_{yx}/t = 200/S_{yx}/t = 10.8 \text{ V K}^{-1}$  for  $\text{Fe}_3\text{Ga}$  ( $S_{yx} \approx 6 \mu\text{V K}^{-1}$ ) and  $7.2 \text{ V K}^{-1}$  for  $\text{Fe}_3\text{Al}$  ( $S_{yx} \approx 4 \mu\text{V K}^{-1}$ ), where  $L$  is the total length of the material. Note that this 'geometrically enhanced' Nernst voltage is comparable to the  $\mu\text{-TEG}$  output for the 25,000 p-n pairs of  $\text{Bi}_2\text{Te}_3$ . For practical applications, the main concern is not widely used efficiency or figure of merit  $ZT$ , but specific power generation capacity  $\Gamma_p = P_{\text{max}}/(A(\Delta T)^2)$ , which indicates how much power is generated from a temperature difference  $\Delta T$  (refs. 11–13).  $\Gamma_p$  for our  $\mu\text{-TEG}$  design is estimated to be  $\Gamma_p = LwS_{yx}^2/(4tp)/(1 \text{ cm}^2) = 324 \mu\text{W (K}^2 \text{ cm}^2)^{-1}$  for  $\text{Fe}_3\text{Ga}$  and  $144 \mu\text{W (K}^2 \text{ cm}^2)^{-1}$  for  $\text{Fe}_3\text{Al}$ . The values estimated here are the hypothetical upper limit. Although the actual values may be reduced by a factor of about ten in a practical  $\mu\text{-TEG}$ , they would be still comparable to or higher than those reported for conventional  $\mu\text{-TEGs}$  to date<sup>11–13</sup>.

### High-throughput calculation of thermoelectric conductivity

For the crystal structures in the Materials Project<sup>28</sup>, the electronic structures were calculated within the generalized gradient approximation using the Quantum Espresso code<sup>42</sup>. The crystal structure was analysed using cif2cell<sup>43</sup> and ultrasoft pseudopotentials in the PSLibrary<sup>44</sup> were used. The magnetic moment was assumed along the  $z$  axis and the SOC's were included. Wannier interpolation was automatically performed by using the wannier90 code<sup>45</sup> with the appropriate initial orbitals and energy ranges; that is, the atomic orbitals whose energy level is higher than  $-1 \text{ Rydberg (Ry)}$  were selected as initial orbitals and the Kohn-Sham orbitals with the energy range below  $E_F$  of  $+1 \text{ eV}$  were fully used to construct the Wannier orbitals. The script used for this calculation is available at <https://github.com/wannier-utils-dev/cif2qewan/>. For each Wannier interpolated Hamiltonian, we confirmed that the energy difference between the interpolated band and the original DFT band is less than 0.2 eV. In terms of the computational cost, we limited the number of atoms in the unit cell to be less than eight atoms. With this criterion, the number of successfully obtained Wannier Hamiltonians is 1,419, including 265 iron-based compounds. The list of  $\alpha_{yx}$  calculated for all the 1,419 compounds are shown in Supplementary Table 1.

For all these Wannier Hamiltonians, the intrinsic contribution of the anomalous Hall conductivity  $\sigma_{xy}$  and the anomalous transverse thermoelectric conductivity  $\alpha_{xy}$  were calculated through the Berry curvature formula as

$$\sigma_{xy}(T, \mu) = -\frac{e^2}{h} \int \frac{d\mathbf{k}}{(2\pi)^3} \Omega_{n,z}(\mathbf{k}) f_{n,\mathbf{k}} \quad (1)$$

$$\alpha_{xy}(T, \mu) = -\frac{1}{e} \int d\epsilon \left( -\frac{\partial f}{\partial \epsilon} \right) \sigma_{xy}(0, \epsilon) \frac{\epsilon - \mu}{T} \quad (2)$$

where  $e$ ,  $\hbar$ ,  $\epsilon$ ,  $f$ ,  $\mu$  and  $\Omega_{n,z}$  are the elementary charge with a negative sign, the reduced Planck constant, the band energy, the Fermi-Dirac

distribution function with the band index  $n$  and the wavevector  $\mathbf{k}$ , the chemical potential and the  $z$  component of the Berry curvature, respectively.

To select the candidates, we first sorted the list by the magnitude of  $\alpha_{\text{max}}$ , which refers to the maximum value of  $|\alpha_{\text{yx}}|$  estimated at the Fermi energy below 500 K. Then, we focused on the iron-based compounds as iron is the least expensive element. Finally, we removed paramagnets, antiferromagnets and the materials that do not exist in past experimental reports.

## DFT calculation

The electronic structures of  $\text{Fe}_3\text{X}$  were obtained by using the OpenMX code<sup>46</sup>, where the exchange-correlation functional within the generalized-gradient approximation and norm-conserving pseudopotentials<sup>47</sup> were employed. The spin-orbit interaction was included by using total angular momentum-dependent pseudopotentials<sup>48</sup>. The wave functions were expanded by a linear combination of multiple pseudoatomic orbitals<sup>49</sup>. A set of pseudoatomic orbital basis was specified as  $\text{Fe}6.0\text{-}s3p3d3$ ,  $\text{Al}7.0\text{-}s3p3d1$  and  $\text{Ga}7.0\text{-}s3p3d3$ , where the number after each element stands for the radial cutoff in the unit of Bohr and the integer after  $s$ ,  $p$  and  $d$  indicates the radial multiplicity of each angular momentum component. The cutoff energies for a charge density of 800.0 Ry and a  $k$ -point mesh of  $35 \times 35 \times 35$  were used. The lattice constants of  $\text{Fe}_3\text{Ga}$  and  $\text{Fe}_3\text{Al}$  were set to 5.80 Å and 5.79 Å, respectively<sup>50,51</sup>. The evaluated magnetic moments of  $\text{Fe}_3\text{Ga}$  and  $\text{Fe}_3\text{Al}$  are 6.69 Bohr magneton per formula unit ( $\mu_{\text{B}}$  f.u.<sup>-1</sup>) and 5.93  $\mu_{\text{B}}$  f.u.<sup>-1</sup>, respectively, which are close to the experimentally obtained values and consistent with previously calculated results<sup>52,53</sup>. The nodal lines were obtained by monitoring the degeneracy of eigenvalues in the momentum space based on the electronic structure without SOC.

## Wannier representation and transport properties

From the Bloch states obtained in the DFT calculation described above, a Wannier basis set was constructed by using the wannier90 code<sup>45</sup>. The basis was composed of  $d$ -character orbitals localized at each Fe site, ( $s, p$ )-character ones at the Al sites and  $p$ -character ones at the Ga sites, that is, 38 orbitals per f.u. of  $\text{Fe}_3\text{Al}$  and 36 orbitals per f.u. of  $\text{Fe}_3\text{Ga}$  in total, including the spin multiplicity. These sets were constructed from the Bloch functions in the energy window ranging from -15 eV to 50 eV, which include 92 bands for  $\text{Fe}_3\text{Al}$  and 110 bands for  $\text{Fe}_3\text{Ga}$ .

The anomalous Hall conductivity  $\sigma_{\text{xy}}$  and the anomalous transverse thermoelectric conductivity  $\alpha_{\text{xy}}$  that intrinsically appear in crystalline systems can be expressed with the out-of-plane component  $\Omega_{n,z}(\mathbf{k})$  of the Berry curvature as described in equations (1) and (2).  $\sigma_{\text{xy}}$  and  $\alpha_{\text{xy}}$  shown in Fig. 3c, d were calculated according to the above equations in the Wannier representation. For the numerical integration, a  $k$ -point mesh of  $100 \times 100 \times 100$  and in addition an adaptive mesh of  $N_{\text{adp}} = 3 \times 3 \times 3$  in regions with large  $\Omega_{n,z}$  were used. This adaptive mesh is already sufficiently large as we confirmed that the relative change of  $-\sigma_{\text{yx}}$  calculated with  $N_{\text{adp}} = 3 \times 3 \times 3$  and  $6 \times 6 \times 6$  is less than 1%.

Extended Data Fig. 5a, b shows the energy dependence of the Hall conductivity  $-\sigma_{\text{yx}}$  and the transverse thermoelectric conductivity estimated using the Mott relation, namely  $\alpha_{\text{yx}}/T = -q(\partial\sigma_{\text{yx}}/\partial E)_{E=E_{\text{F}}}$  with  $q = \frac{\pi^2 k_{\text{B}}^2}{3|e|}$ . Extended Data Fig. 5c shows the temperature dependence of  $-\alpha_{\text{yx}}/T$  calculated using equation (2) for various energies.

## Data availability

The data that support the plots within this paper and the other findings of this study are available from the corresponding author upon reasonable request.

39. Slack, G. A. Thermal conductivity of  $\text{MgO}$ ,  $\text{Al}_2\text{O}_3$ ,  $\text{MgAl}_2\text{O}_4$ , and  $\text{Fe}_3\text{O}_4$  crystals from 3° to 300° K. *Phys. Rev.* **126**, 427–441 (1962).
40. Rudajevová, A. & Buriánek, J. Determination of thermal diffusivity and thermal conductivity of Fe-Al alloys in the concentration range 22 to 50 at. % Al. *J. Phase Equilibria* **22**, 560–563 (2001).
41. Gaviot, E. et al. Distribution-patterned radiometers: a new paradigm for irradiance measurement. *Proc. SPIE* **3061**, 800–810 (1997).
42. Giannozzi, P. et al. QUANTUM ESPRESSO: a modular and open-source software project for quantum simulations of materials. *J. Phys. Condens. Matter* **21**, 395502 (2009).
43. Björkman, T. CIF2Cell: generating geometries for electronic structure programs. *Comput. Phys. Commun.* **182**, 1183–1186 (2011).
44. Dal Corso, A. Pseudopotentials periodic table: from H to Pu. *Comput. Mater. Sci.* **95**, 337–350 (2014).
45. Mostofi, A. A. et al. An updated version of wannier90: a tool for obtaining maximally localised Wannier functions. *Comput. Phys. Commun.* **185**, 2309–2310 (2014).
46. Ozaki, T. et al. OpenMX: open source package for Material eXplorer (2019); <http://www.openmx-square.org/>
47. Morrison, I., Bylander, D. M. & Kleinman, L. Nonlocal Hermitian norm-conserving Vanderbilt pseudopotential. *Phys. Rev. B* **47**, 6728–6731 (1993).
48. Theurich, G. & Hill, N. A. Self-consistent treatment of spin-orbit coupling in solids using relativistic fully separable ab initio pseudopotentials. *Phys. Rev. B* **64**, 073106 (2001).
49. Ozaki, T. Variationally optimized atomic orbitals for large-scale electronic structures. *Phys. Rev. B* **67**, 155108 (2003).
50. Nishino, Y. et al. Semiconductor like behavior of electrical resistivity in Heusler-type  $\text{Fe}_2\text{VAl}$  compound. *Phys. Rev. Lett.* **79**, 1909–1912 (1997).
51. Matyunina, M., Zagrebina, M., Sokolovskiy, V. & Buchelnikov, V. Ab initio study of magnetic and structural properties of Fe-Ga alloys. *EPJ Web Conf.* **185**, 04013 (2018).
52. Lechermann, F. et al. Density-functional study of  $\text{Fe}_3\text{Al}$ : LSDA versus GGA. *Phys. Rev. B* **65**, 132104 (2002).
53. Paduani, C. & Bormio-Nunes, C. Density functional theory study of  $\text{Fe}_3\text{Ga}$ . *J. Appl. Phys.* **109**, 033705 (2011).

**Acknowledgements** We thank T. Tsujikawa for assistance with thin-film fabrication. This work is partially supported by CREST (JPMJCR18T3), New Energy and Industrial Technology Development Organization (NEDO), PRESTO (JPMJPR15N5), Japan Science and Technology Agency, by Grants-in-Aids for Scientific Research on Innovative Areas (JP15H05882 and JP15H05883) from the Ministry of Education, Culture, Sports, Science, and Technology of Japan, and by Grants-in-Aid for Scientific Research (JP16H02209, JP16H06345, JP19H00650) from the Japanese Society for the Promotion of Science (JSPS). The work at IQM was supported by the US Department of Energy, Office of Basic Energy Sciences, Division of Materials Sciences and Engineering under grant DE-FG02-08ER46544. The work for first-principles calculations was supported in part by JSPS Grant-in-Aid for Scientific Research on Innovative Areas (JP18H04481 and JP19H05825) and by MEXT as a social and scientific priority issue (Creation of new functional devices and high-performance materials to support next-generation industries) to be tackled by using post-K computer (hp180206 and hp190169). The use of the facilities of the Materials Design and Characterization Laboratory at the Institute for Solid State Physics, The University of Tokyo, is acknowledged.

**Author contributions** A.S., S. Minami, T.K., T.C. and T.H. contributed equally to this work. S.N. and R.A. conceived the project. S.N. planned the experiments. T.K. performed the high-throughput computational search. T.C. and Y.W. worked on the single-crystal growth and the preparation of samples. A.S., T.C. and Y.W. carried out the transport and magnetization measurements and analysed the data. T.H. and S. Miwa fabricated the thin film and performed its structural and chemical analyses and transport measurement. S. Minami, T.K., F.I., T.N., M.H. and R.A. performed the first-principles calculations. D.S.-H. performed chemical analyses and took the electron diffraction image. S.N., R.A., A.S., S. Minami, T.K., T.H. and F.I. wrote the paper. All authors discussed the results and commented on the manuscript.

**Competing interests** The authors declare no competing interests.

## Additional information

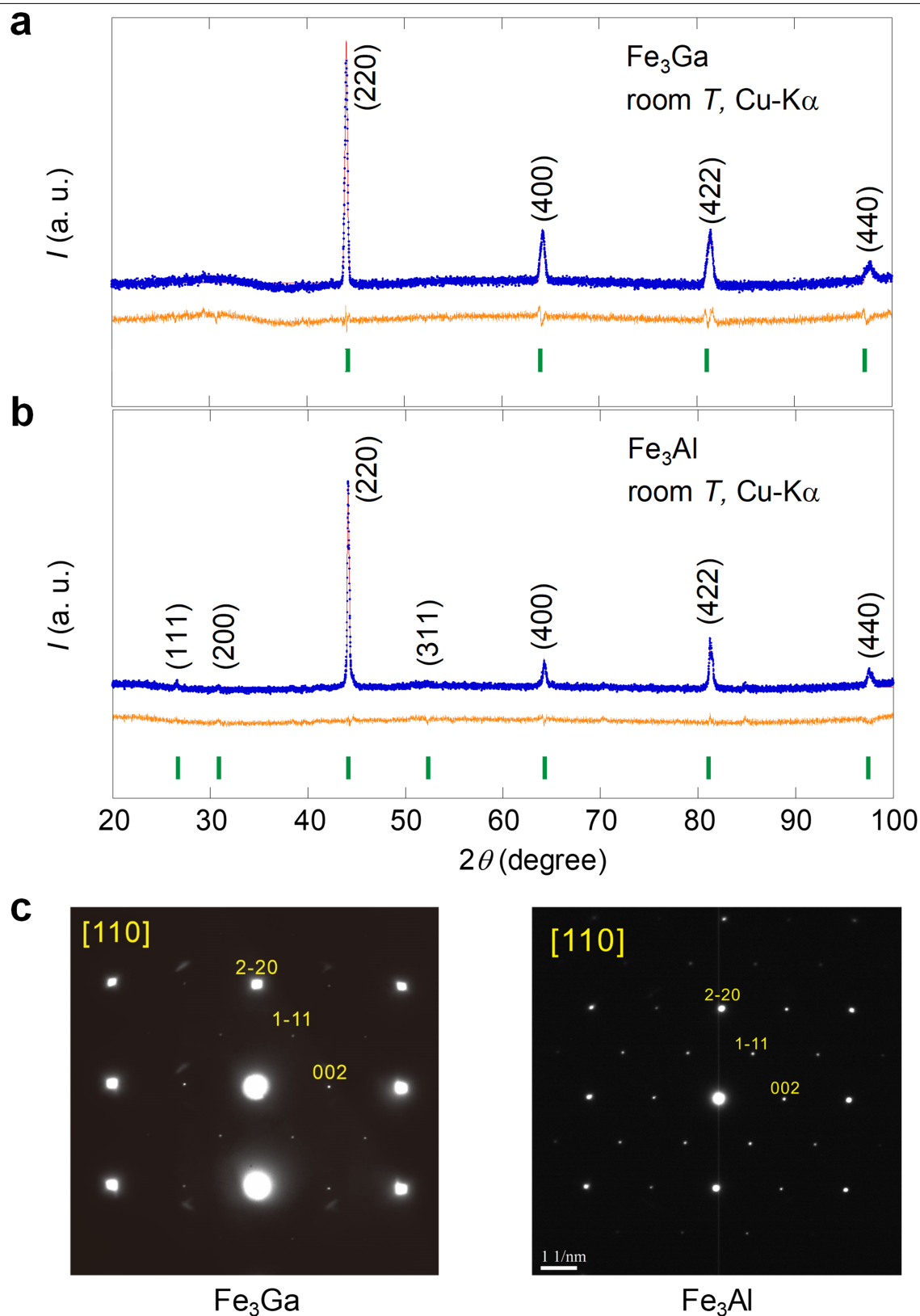
**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41586-020-2230-z>.

**Correspondence and requests for materials** should be addressed to S.N.

**Peer review information** Nature thanks Ernst Bauer and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

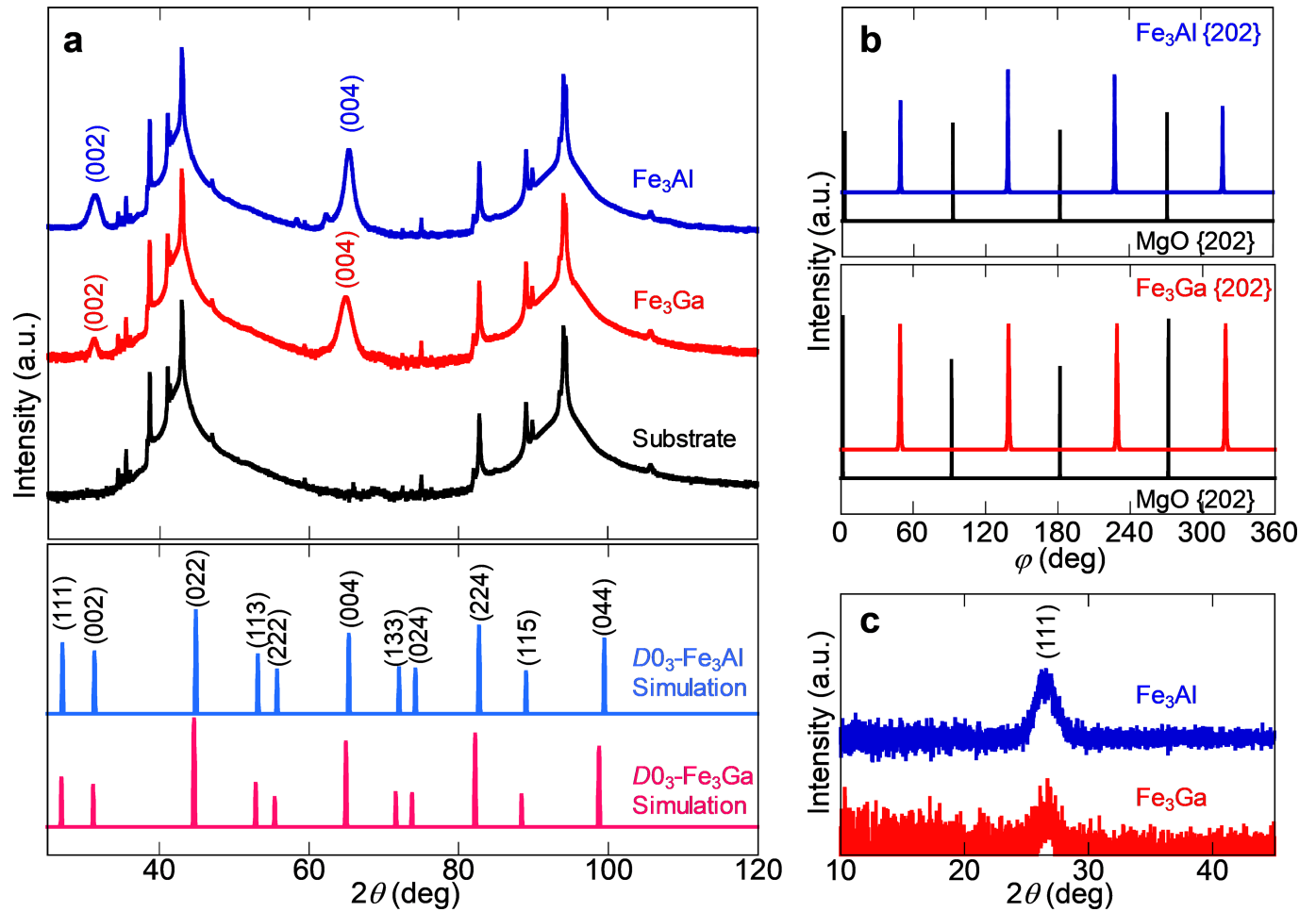
**Reprints and permissions information** is available at <http://www.nature.com/reprints>.





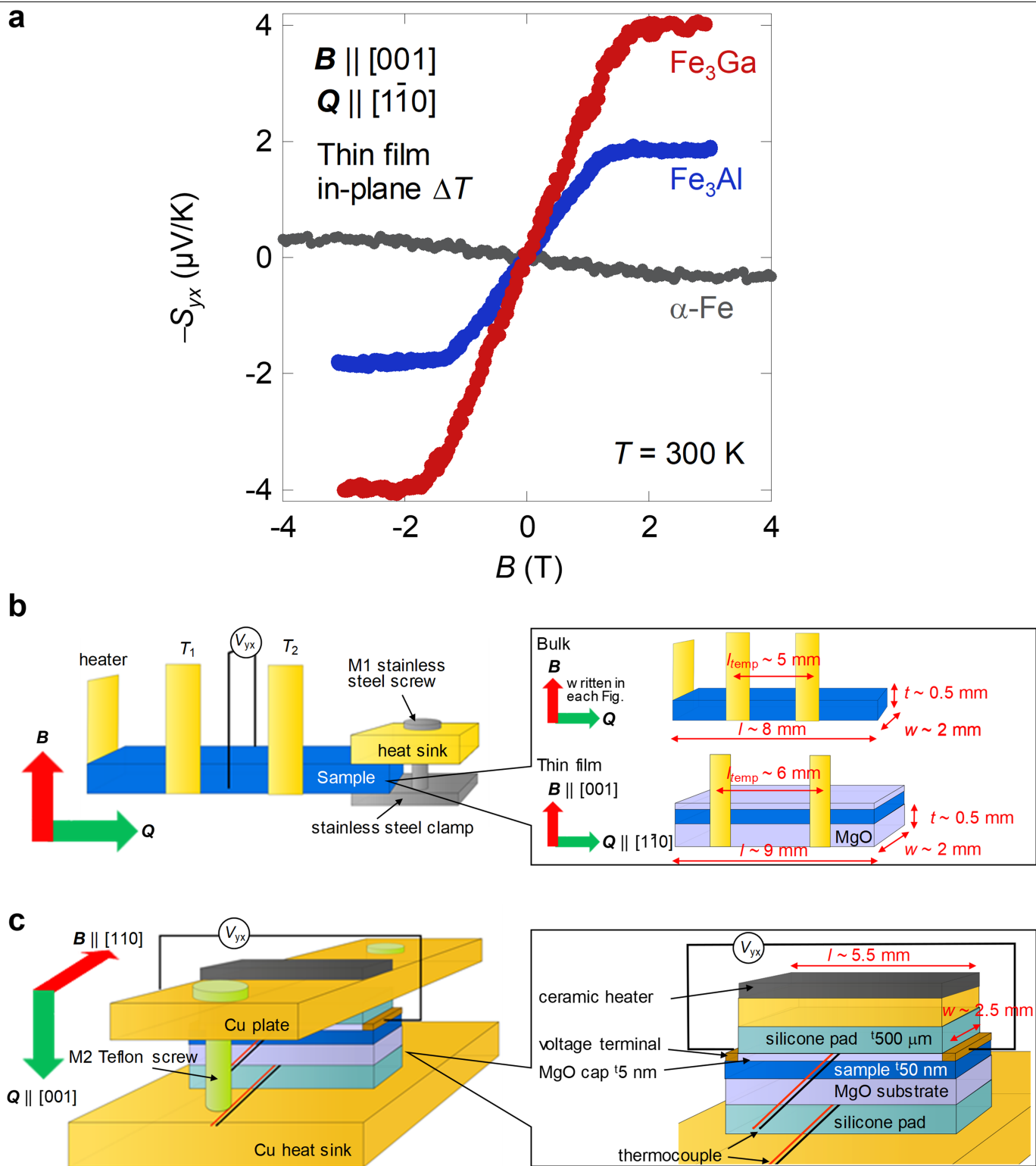
**Extended Data Fig. 1 | Evidence for the  $D0_3$  structure of  $\text{Fe}_3\text{X}$ .** **a, b**, XRD pattern for  $\text{Fe}_3\text{Ga}$  (**a**) and  $\text{Fe}_3\text{Al}$  (**b**) obtained by Cu-Kα radiation ( $\lambda = 1.5401 \text{ \AA}$ ) at room temperature. The solid circles and the solid line (red) represent the experimental results and the Rietveld refinement fit, respectively. The final weighted and expected  $R$  indicators and goodness-of-fit indicator  $S$  are  $R_{\text{WP}} = 2.34\%$ ,  $R_e = 1.49\%$  and  $S = 1.55$  for  $\text{Fe}_3\text{Ga}$  and  $R_{\text{WP}} = 1.91\%$ ,  $R_e = 1.27\%$  and

$S = 1.48$  for  $\text{Fe}_3\text{Al}$ , respectively. Vertical bars (green) below the curves indicate the major peak positions calculated for  $D0_3$   $\text{Fe}_3\text{Ga}$  and  $\text{Fe}_3\text{Al}$ , which are more than 1% of the main peak. The lower curve (orange) represents the difference between the experimental result and the Rietveld refinement. *I*, intensity. **c**, Selected area electron diffraction pattern for our single crystals of  $\text{Fe}_3\text{Ga}$  (left) and  $\text{Fe}_3\text{Al}$  (right) taken from the [110] plane.



**Extended Data Fig. 2 | Evidence for  $D0_3$  structure of  $\text{Fe}_3\text{Ga}$  and  $\text{Fe}_3\text{Al}$  thin films.** **a**, Room temperature spectra obtained by XRD  $2\theta/\omega$ -scans for the  $\text{Fe}_3\text{X}$  thin films on an MgO substrate and the MgO substrate itself. The theoretical simulation patterns for the  $D0_3$   $\text{Fe}_3\text{Ga}$  and  $\text{Fe}_3\text{Al}$  structures are presented at the

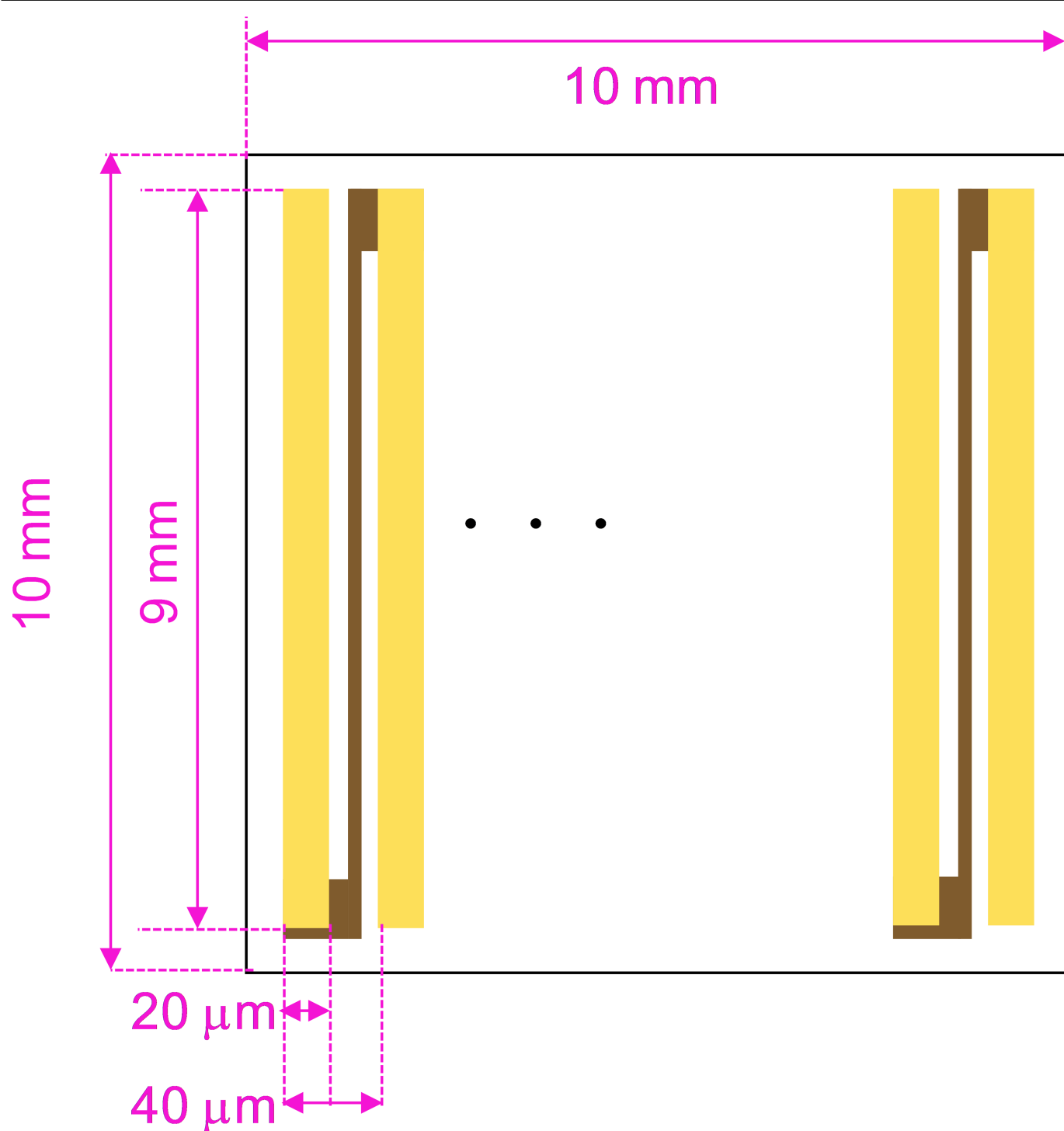
bottom. **b**,  $\phi$ -scan patterns of the  $\{202\}$  planes of the  $\text{Fe}_3\text{Ga}$  and  $\text{Fe}_3\text{Al}$  layers, and the MgO substrate. **c**,  $2\theta/\omega$ -scan patterns for the  $(111)$  plane of the  $\text{Fe}_3\text{Ga}$  and  $\text{Fe}_3\text{Al}$  thin-film layers.



**Extended Data Fig. 3 | ANE of  $\text{Fe}_3\text{X}$  and experimental setup for both bulk and thin films. a,** Magnetic field dependence of the ANE obtained for the  $\text{Fe}_3\text{X}$  thin films (50 nm) and the  $\alpha\text{-Fe}$  thin film (50 nm) using the in-plane temperature

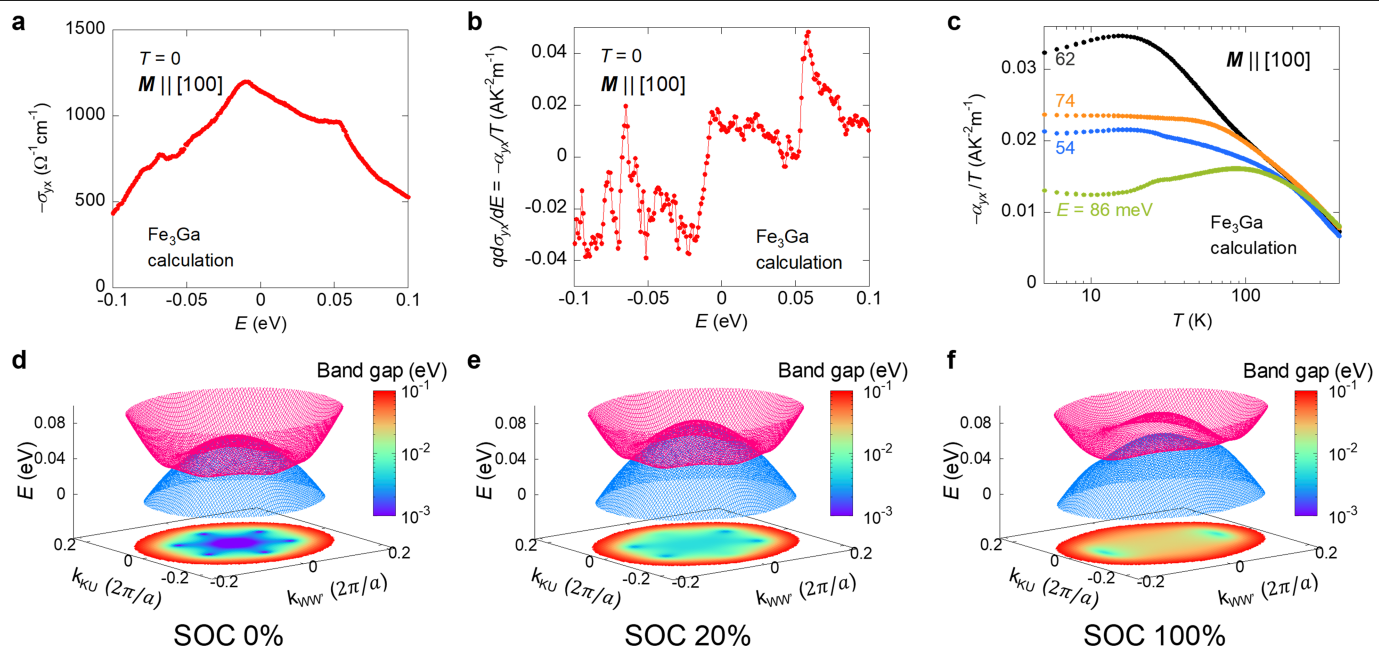
gradient. **b,** Schematic of the experimental setup for the ANE measurement using the in-plane temperature gradient. **c,** Schematic of the experimental setup for the ANE measurement using the out-of-plane temperature gradient.





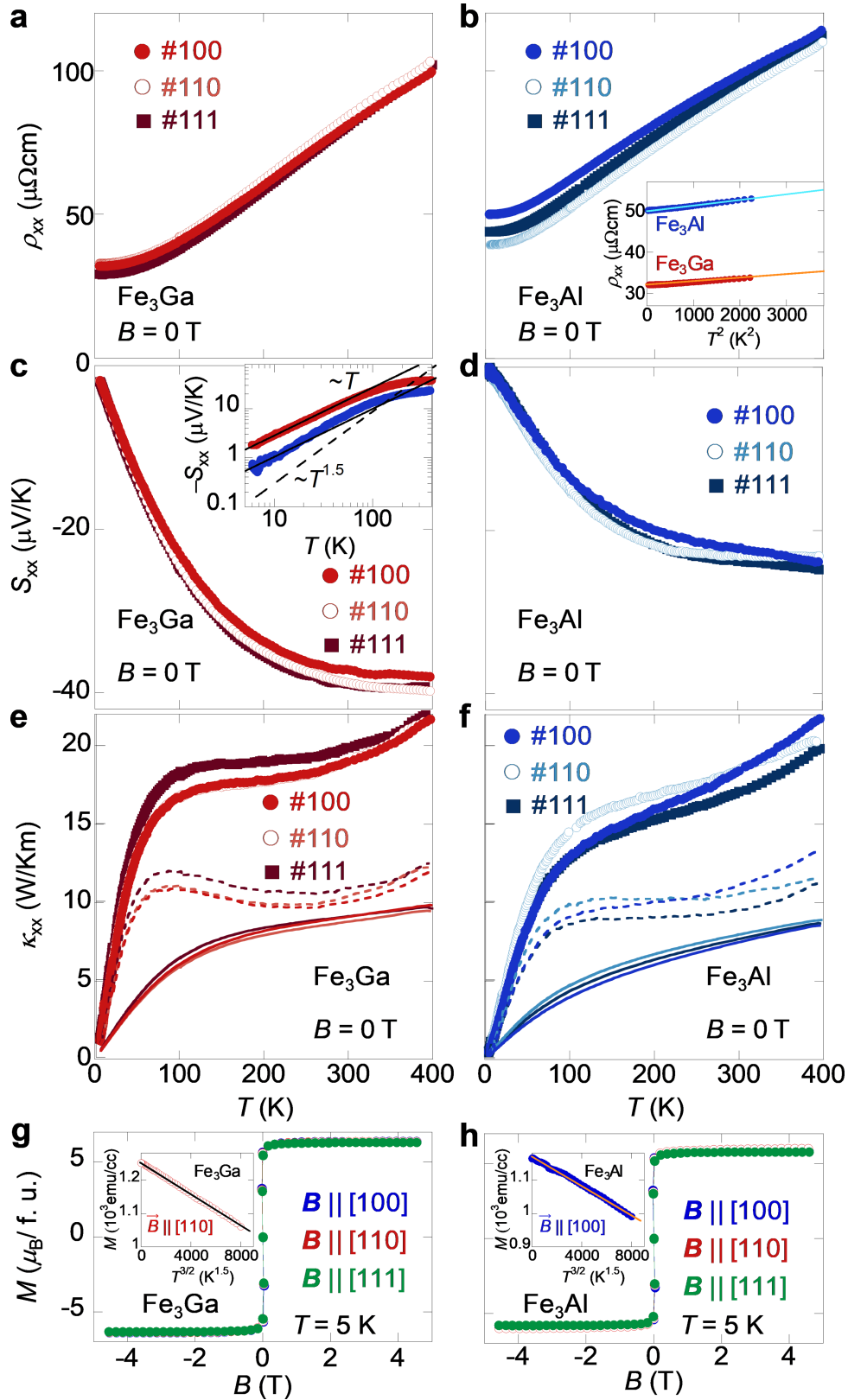
**Extended Data Fig. 4 | Schematic of the  $\mu$ -TEG based on the ANE.** The thermopile consists of a square-shaped substrate (black frame) and an alternating array of Fe<sub>3</sub>X (yellow) and gold wires (brown) placed on the

substrate and these two wires are connected in a zigzag configuration. A temperature gradient is applied perpendicular to the plane. The thickness of the wire is designed to be 1  $\mu$ m.



**Extended Data Fig. 5 | Energy dependence of the Hall conductivity and transverse thermoelectric conductivity, and the effect of the SOC on the nodal web structure.** **a**, Energy dependence of  $-\sigma_{yx}$  obtained from the first-principles calculations at  $T = 0$ . **b**, Energy dependence of  $-\alpha_{yx}/T$  calculated

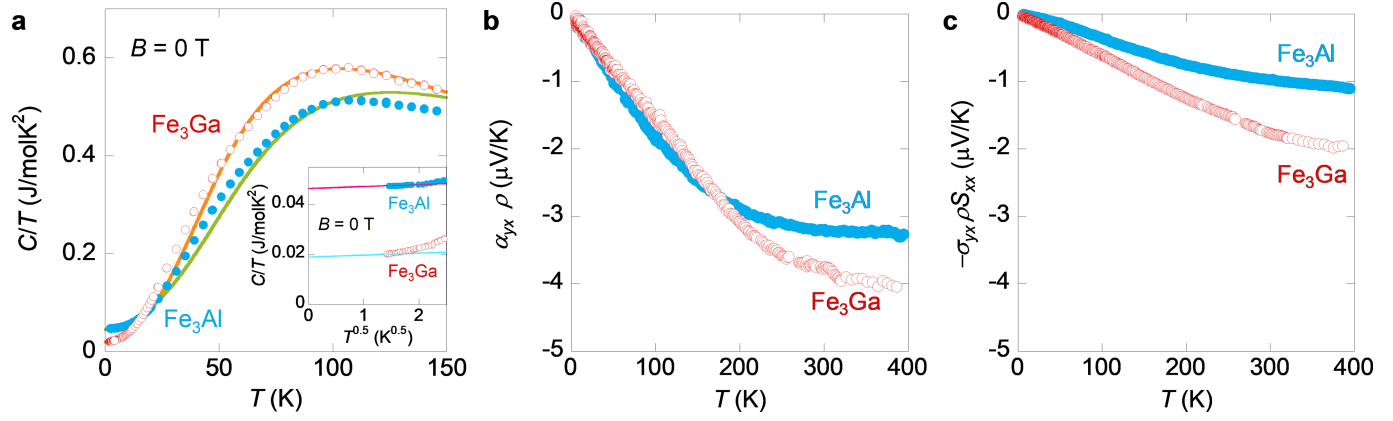
based on the Mott relation (Methods). **c**, Temperature dependence of  $-\alpha_{yx}/T$  at various energies. **d–f**, Band structure of the nodal web around the L point for different strengths of the SOC: 0% (**d**), 20% (**e**) and 100% (**f**). We also show the contour plot of the bandgap in the  $k_{KU}$ – $k_{WW}$  plane.



**Extended Data Fig. 6 | Longitudinal electric and thermal transport properties and magnetization for  $\text{Fe}_3\text{X}$ .** **a, b**, Temperature dependence of the longitudinal resistivity  $\rho_{xx}$  for  $\text{Fe}_3\text{Ga}$  (**a**) and  $\text{Fe}_3\text{Al}$  (**b**). **c, d**, Temperature dependence of the Seebeck coefficient  $S_{xx}$  for  $\text{Fe}_3\text{Ga}$  (**c**) and  $\text{Fe}_3\text{Al}$  (**d**). **e, f**, Temperature dependence of the thermal conductivity  $\kappa_{xx}$  for  $\text{Fe}_3\text{Ga}$  (**e**) and  $\text{Fe}_3\text{Al}$  (**f**). #100, #110 and #111 represent the samples used for the transport measurements in  $B \parallel [100]$ ,  $[110]$  and  $[111]$ , respectively. The inset in **b** shows the  $T^2$  dependence of  $\rho_{xx}$  for  $\text{Fe}_3\text{Ga}$  #100 (red) and  $\text{Fe}_3\text{Al}$  #100 (blue). The inset in **c**

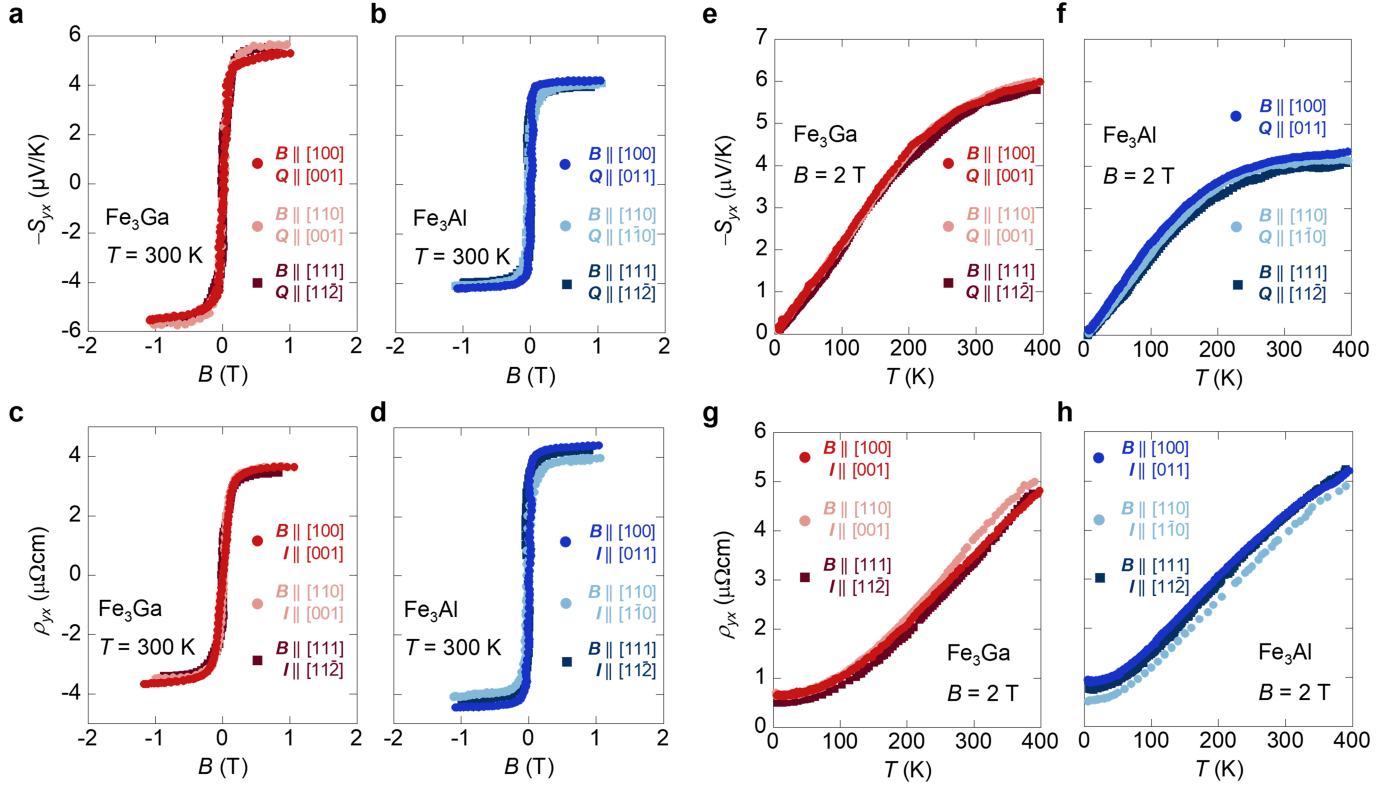
shows the log-log plot of  $-S_{xx}$  versus  $T$  for  $\text{Fe}_3\text{Ga}$  #100 (red) and  $\text{Fe}_3\text{Al}$  #100 (blue). The solid and broken lines represent the  $T$  and  $T^{1.5}$  dependence, respectively. The solid and broken lines in **e** and **f** show the estimated electric and lattice contributions to the thermal conductivity (Supplementary Information). **g, h**, Magnetization curve for  $\text{Fe}_3\text{Ga}$  (**g**) and  $\text{Fe}_3\text{Al}$  (**h**) at  $T = 5$  K under  $B \parallel [100]$ ,  $[110]$  and  $[111]$ . The insets in **g** and **h** are the  $T^{3/2}$  dependence of  $M(T)$ .





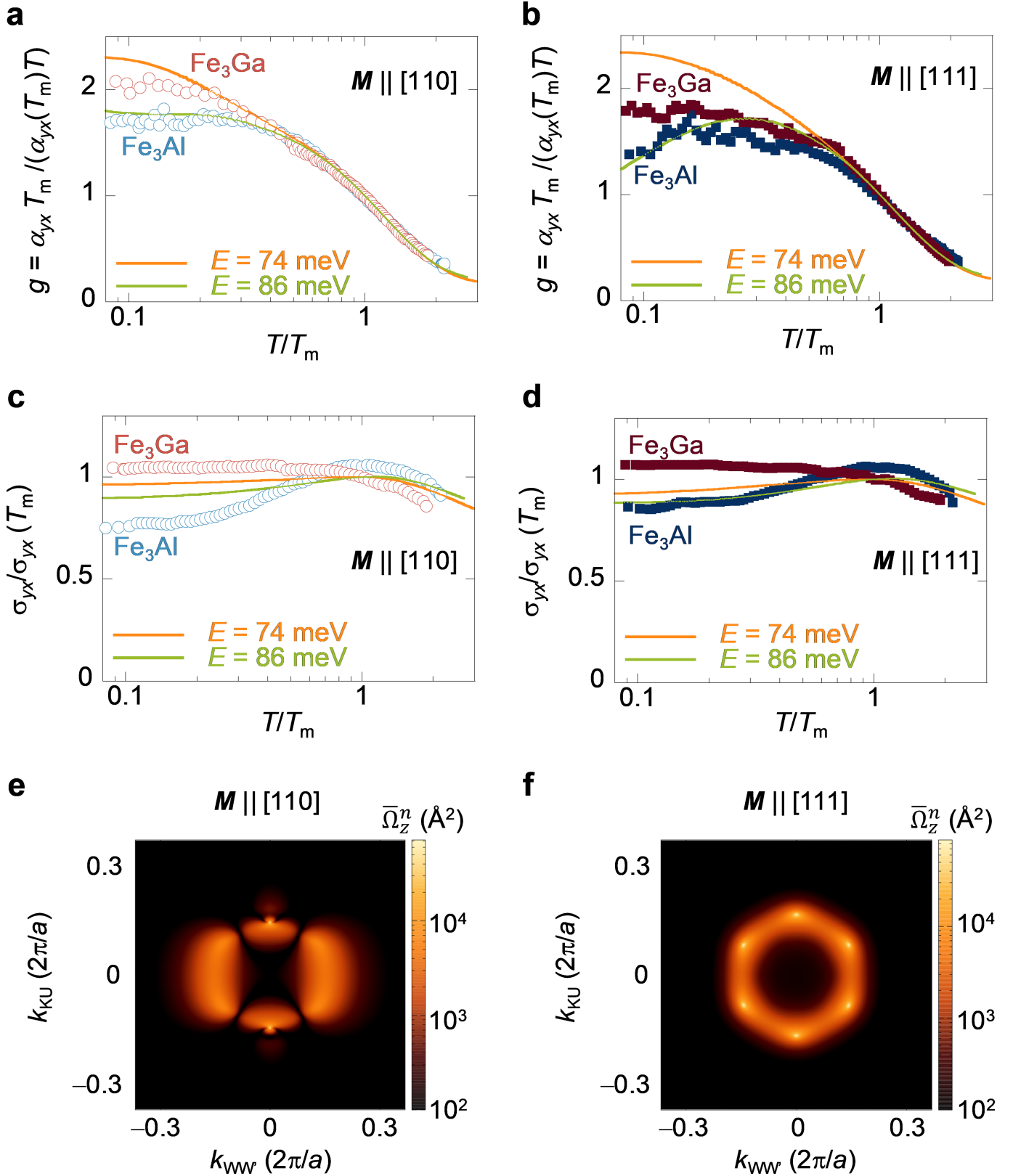
**Extended Data Fig. 7 | Specific heat and two contributions to the anomalous Nernst effect.** **a**, Temperature dependence of the specific heat divided by temperature  $C/T$  for  $\text{Fe}_3\text{X}$ . The solid lines represent the fit by the combination of the electronic and Debye-type phonon specific heat and

ferromagnetic magnon contribution (Supplementary Information). The inset shows the  $T^{1/2}$  dependence of  $C/T$  at low temperatures. **b**, **c**,  $\alpha_{yx} \rho$  (**b**) and  $-\sigma_{yx} \rho S_{xx}$  (**c**). The ANE is the sum of the two terms, that is,  $S_{yx} = \alpha_{yx} \rho - \sigma_{yx} \rho S_{xx}$  (Supplementary Information).



**Extended Data Fig. 8 | Anisotropy in the Nernst coefficient and Hall resistivity. a–d,** Magnetic field dependence of the Nernst coefficient  $-S_{yx}$  for Fe<sub>3</sub>Ga (a) and Fe<sub>3</sub>Al (b), and the Hall resistivity  $\rho_{yx}$  for Fe<sub>3</sub>Ga (c) and Fe<sub>3</sub>Al (d) in

$B \parallel [100]$ ,  $[110]$  and  $[111]$ . **e–h** Temperature dependence of  $-S_{yx}$  for Fe<sub>3</sub>Ga (e) and Fe<sub>3</sub>Al (f), and  $\rho_{yx}$  for Fe<sub>3</sub>Ga (g) and Fe<sub>3</sub>Al (h) in  $B \parallel [100]$ ,  $[110]$  and  $[111]$ .



**Extended Data Fig. 9 | Anisotropy in the transverse thermoelectric conductivity  $-\alpha_{yx}$ , Hall conductivity  $-\sigma_{yx}$  and Berry curvature. a, b, Scaling relation of  $-\alpha_{yx}$  for  $M \parallel [110]$  (a) and  $M \parallel [111]$  (b) versus  $T/T_m$ . c, d, Scaling relation of  $-\sigma_{yx}$  for  $M \parallel [110]$  (c) and  $M \parallel [111]$  (d) versus  $T/T_m$ . The solid lines in a–d are obtained by the first-principles calculations. The scaling parameters used here**

are summarized in Extended Data Table 1. Details are the same as Fig. 3c, d for  $M \parallel [100]$  in the main text. e, f, Contour plot of the Berry curvature  $\Omega_{n,z}$  of the lower-energy band  $n$  in the vicinity of the nodal web structure around the L point for  $M \parallel [110]$  (e) and  $M \parallel [111]$  (f). Details are the same as Fig. 4e for  $M \parallel [100]$  in the main text.

Extended Data Table 1 | List of the scaling parameters and magnitude of  $\sigma_{yx}$  and  $\sigma_{xx}$  at  $T \approx 0$

	<i>M</i>	<i>T</i> <sub>m</sub>	$-\alpha_{yx}(T_m)$	$-\sigma_{yx}(T_m)$	$ \sigma_{yx}(T \sim 0) $		$ \sigma_{xx}(T \sim 0) $	
unit	-	K	AK <sup>-1</sup> m <sup>-1</sup>	Ω <sup>-1</sup> cm <sup>-1</sup>	Ω <sup>-1</sup> cm <sup>-1</sup>	<i>e</i> <sup>2</sup> /(2 <i>h</i> <i>a</i> )	Ω <sup>-1</sup> cm <sup>-1</sup>	<i>e</i> <sup>2</sup> /(2 <i>h</i> <i>a</i> )
Fe <sub>3</sub> Ga	[100]	200	5.2	610	~600	~0.9	~31000	~47
	[110]	210	5	570				
	[111]	200	5.1	550				
Fe <sub>3</sub> Al	[100]	180	3.6	460	~370	~0.55	~20000	~30
	[110]	180	3.8	460				
	[111]	180	3.3	460				
74 meV	[100]	320	3.3	730				
	[110]	340	2.8	610				
	[111]	340	2.7	590				
86 meV	[100]	370	3.3	670				
	[110]	370	2.8	560				
	[111]	380	2.7	540				

The first two rows indicate the experimental values obtained for Fe<sub>3</sub>Ga and Fe<sub>3</sub>Al and the last three rows the theoretical values obtained for *E* = +74 and *E* = +86 meV. Each has three consecutive rows corresponding to the orientation of the magnetization. Here, *T*<sub>m</sub> is the temperature where the transverse thermoelectric conductivity  $|\alpha_{yx}|$  attains its maximum  $|\alpha_{yx}(T_m)|$ .



# Liquid flow and control without solid walls

<https://doi.org/10.1038/s41586-020-2254-4>

Received: 1 July 2018

Accepted: 26 February 2020

Published online: 6 May 2020



Peter Dunne<sup>1,2,8</sup>, Takuji Adachi<sup>1,3,8</sup>, Arvind Arun Dev<sup>2</sup>, Alessandro Sorrenti<sup>1,4</sup>, Lucas Giacchetti<sup>1</sup>, Anne Bonnin<sup>5</sup>, Catherine Bourdon<sup>6</sup>, Pierre H. Mangin<sup>6</sup>, J. M. D. Coey<sup>7</sup>, Bernard Doudin<sup>2</sup> & Thomas M. Hermans<sup>1✉</sup>

When miniaturizing fluidic circuitry, the solid walls of the fluid channels become increasingly important<sup>1</sup> because they limit the flow rates achievable for a given pressure drop, and they are prone to fouling<sup>2</sup>. Approaches for reducing the wall interactions include hydrophobic coatings<sup>3</sup>, liquid-infused porous surfaces<sup>4–6</sup>, nanoparticle surfactant jamming<sup>7</sup>, changes to surface electronic structure<sup>8</sup>, electrowetting<sup>9,10</sup>, surface tension pinning<sup>11,12</sup> and use of atomically flat channels<sup>13</sup>. A better solution may be to avoid the solid walls altogether. Droplet microfluidics and sheath flow achieve this but require continuous flow of the central liquid and the surrounding liquid<sup>1,14</sup>. Here we demonstrate an approach in which aqueous liquid channels are surrounded by an immiscible magnetic liquid, both of which are stabilized by a quadrupolar magnetic field. This creates self-healing, non-clogging, anti-fouling and near-frictionless liquid-in-liquid fluidic channels. Manipulation of the field provides flow control, such as valving, splitting, merging and pumping. The latter is achieved by moving permanent magnets that have no physical contact with the liquid channel. We show that this magnetostatic pumping method can be used to transport whole human blood with very little damage due to shear forces. Haemolysis (rupture of blood cells) is reduced by an order of magnitude compared with traditional peristaltic pumping, in which blood is mechanically squeezed through a plastic tube. Our liquid-in-liquid approach provides new ways to transport delicate liquids, particularly when scaling channels down to the micrometre scale, with no need for high pressures, and could also be used for microfluidic circuitry.

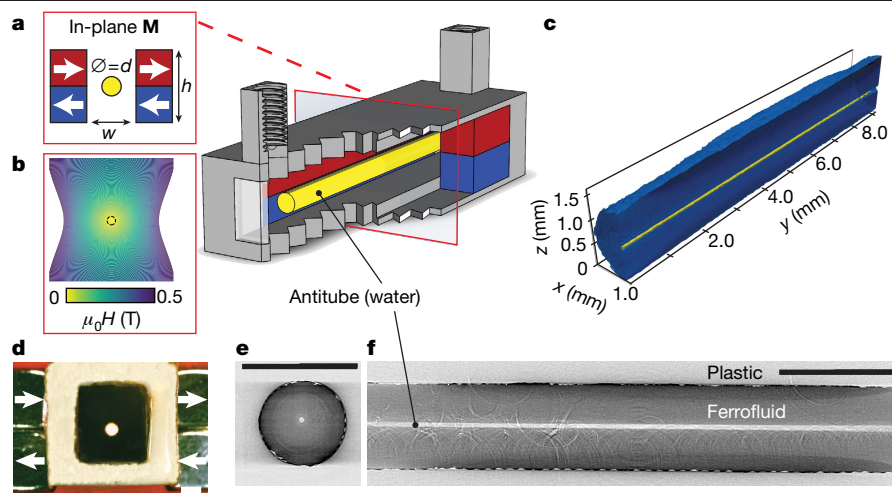
To avoid contact with the walls of a device, magnetic forces can be used to levitate particles or live cells in a paramagnetic liquid<sup>15</sup>. Nearly wall-less microfluidic channels have been demonstrated in which continuous ‘magnetic antitubes’ of water surrounded by an aqueous paramagnetic salt solution were stabilized<sup>16</sup> by using an electromagnet and iron tracks. However, the antitube lifetime was limited by ion interdiffusion between the two liquids, the salts were toxic, and contact with one stationary wall containing the track could not be avoided. Here we overcome these limitations, creating entirely wall-less microfluidic channels consisting of diamagnetic antitubes completely enclosed by immiscible, and in select cases, non-toxic, paramagnetic fluids.

The key to wall-less magnetic confinement is an extended quadrupolar flux source, leading to a null magnetic field along a line at the centre (Fig. 1a, b). Appropriately magnetized Nd<sub>2</sub>Fe<sub>14</sub>B bars are used to define tubular channels, or else custom-made magnet bilayers are waterjet-cut to define more complex fluidic circuitry. The magnetic bars are housed in 3D-printed supports with conventional microfluidic inlet and outlet ports (Fig. 1 and Extended Data Fig. 1). The various magnetic fluids used are listed in Table 1; stronger confinement is achieved with ferrofluids, but optical transparency is possible with our paramagnetic oil, which we call ‘Magoil’. Ferrofluids are colloidal liquids of magnetite (Fe<sub>3</sub>O<sub>4</sub>) nanoparticles suspended in a carrier fluid, whereas Magoil is a rare-earth-based oil,

inspired by the contrast agents based on diethylenetriaminepentaacetate (DTPA) that are used for magnetic resonance imaging<sup>17</sup> (Extended Data Fig. 2, and Methods). We used commercial ferrofluids (Ferrotec APG311 and EMG900, 2 vol% and 13 vol% magnetite in hydrocarbon oils respectively; and Qfluidics MKC and MD4, 29 vol% magnetite in hydrocarbons and 5 vol% magnetite in perfluorodecalin, respectively).

Aqueous antitubes were typically formed by injecting ferrofluid into water-filled quadrupole channels and were visualized by X-ray imaging and tomography (Fig. 1c, d, f, Extended Data Fig. 3c–e). Tomography images unambiguously reveal the wall-less features, confirmed by visual inspection of transmitted light through antitubes up to 1 m long (Extended Data Fig. 1g). Optical imaging was found to be possible through ferrofluid no more than 200 μm thick and with a high-contrast camera (Extended Data Fig. 3f, g), where we can take advantage of the optical resolution to image the smallest antitube features. We were able to image antitubes in transparent Magoil with standard optical or fluorescent microscopy by adding contrast ink or fluorescent dye to the water antitube, allowing real-time visualization of antitube extrusion and retraction (Supplementary Video 1). Furthermore, the trapped gas bubbles that are often problematic in conventional devices can easily be removed, as their buoyancy in Magoil overcomes the magnetic confinement. For most practical applications, however, ferrofluids are preferred because of their much

<sup>1</sup>Université de Strasbourg, CNRS, ISIS, Strasbourg, France. <sup>2</sup>Université de Strasbourg, CNRS, IPCMS UMR 7504, Strasbourg, France. <sup>3</sup>Department of Physical Chemistry, University of Geneva, Geneva, Switzerland. <sup>4</sup>Institute for Chemical and Bioengineering, Department of Chemistry and Applied Biosciences, ETH Zürich, Zurich, Switzerland. <sup>5</sup>Paul Scherrer Institut, Villigen, Switzerland. <sup>6</sup>Université de Strasbourg, INSERM, EFS Grand-Est, BPPS UMR-S1255, FMTS, Strasbourg, France. <sup>7</sup>School of Physics and CRANN, Trinity College, Dublin, Ireland. <sup>8</sup>These authors contributed equally: Peter Dunne, Takuji Adachi. ✉e-mail: [hermans@unistra.fr](mailto:hermans@unistra.fr)



**Fig. 1 | Wall-less magnetic confinement in a fluidic channel.** **a**, Permanent magnets (red, blue) in an in-plane quadrupolar configuration create a low-field zone at the centre, where an antitube of water (yellow) is stabilized inside an immiscible magnetic liquid (white). **b**, Contour plot of the magnetic field. **c**, Synchrotron X-ray tomographic reconstruction of a water antitube (yellow)

larger magnetic susceptibility  $\chi$  (Table 1); they can withstand considerable flow rates while remaining confined in the quadrupole (Extended Data Fig. 4a, b). A 1-mm-diameter liquid tube can deliver a flow of about  $40 \text{ ml min}^{-1}$ . In addition, the phase transfer of magnetite nanoparticles into water is very low, with measured iron concentrations mostly below 1 ppm (see Methods section ‘Quantitative analysis’).

The liquid-in-liquid design offers advantages of stability and robustness for fluid transport. Figure 2a illustrates self-healing after an antitube in ferrofluid was severed with a spatula. Recovery without applied external pressure is rapid (self-healing in Magoil is illustrated in Supplementary Video 2). The antitubes cannot clog: when glass beads are introduced (Fig. 2b), they are easily flushed out. Even a bead much larger than the antitube diameter can be pushed through using minimal pressure (20 mbar; Fig. 2c). The liquid walls of the antitube stretch to avoid clogging and return to their original size when the obstruction is expelled. A change of external pressure alters the antitube diameter. In Extended Data Fig. 4c, d, it can be seen that antitubes remain unchanged with externally applied pressure for an open outlet at atmospheric pressure, but dilate when the same pressure is applied with the outlet closed off. This can also be seen in Fig. 2c, where the tube dilates behind the bead to accommodate the increased local pressure. As a demonstration of anti-fouling behaviour, we used laser light to covalently crosslink a polymer inside an antitube, and were able to remove the resulting solid polymer rod from the outlet of the device (Extended Data Fig. 6a–c, and Supplementary Video 3).

A further advantage of liquid-in-liquid flow is near-frictionless transport with negligible pressure drop. An illustration is shown in Fig. 2d and Supplementary Video 4, where flow of a magnetically confined antitube made of honey (dynamic viscosity  $\eta_h = 10 \text{ Pa s}$ ) is compared with honey flow in a standard tube. We observed an antitube flow of  $39.4 \pm 0.7 \text{ g h}^{-1}$ , about 70 times faster than the flow through a conventional plastic tube of the same diameter  $d = 1.1 \text{ mm}$  ( $0.55 \pm 0.10 \text{ g h}^{-1}$ ). The friction inside a normal tube is quantified by its slip length, defined as the extrapolated distance relative to the boundary where the velocity reduces to zero. Here, the ferrofluid acts as a lubricating layer, with an effective slip length  $b$  at the honey/ferrofluid boundary that can be approximated by

$$b = \frac{1}{2} \left[ \sqrt{\frac{d^2}{2} (Q_{\text{ND}} + 1)} - d \right] \quad (1)$$

with diameter  $d$  is the antitube diameter and  $Q_{\text{ND}}$  is the non-dimensional flow rate given by

$$Q_{\text{ND}} = \frac{128Q\eta_h}{\pi d^4 \partial P / \partial z} \quad (2)$$

where  $Q$  is the flow rate and  $\partial P / \partial z$  is the pressure gradient in the flow direction (for derivation, see Methods section ‘Derivation of equations’). Experimentally, we obtain an effective slip length of 4.3 mm, meaning that under these conditions the flow is essentially plug-like. Under the assumption of an infinite channel, the plug-like velocity profile can be calculated, as shown analytically in Extended Data Fig. 4e and using numerical methods in Extended Data Fig. 5a–d, leading to a very long theoretical effective slip length of 8.1 mm. Remarkably, the flow rate of honey through the antitube was 1.5 times as fast as when it fell freely and unconfined (Fig. 2d, centre), probably owing to competition between orifice wetting<sup>18</sup> and the higher hydrostatic pressure due to the greater height of the honey column in the antitube design.

The stable confinement of an antitube, at equilibrium, results from the competing magnetic energy of the confining fluid and the surface energy  $\sigma$  of the magnetic/non-magnetic interface. Inserting these energy densities into the magnetically augmented version of Bernoulli’s equation<sup>19</sup> gives the equilibrium diameter of the antitube (see Methods section ‘Derivation of equilibrium diameter equation’):

$$d = \frac{4\sigma}{2\mu_0 \bar{M} H_1 + \mu_0 M_1^2} \quad (3)$$

where  $H_1$ ,  $M_1$  are the magnetic field and magnetization values at the interface,  $\mu_0$  is the permeability of free space, and  $\bar{M}$  is the field-averaged magnetization of the confining fluid. This equation is valid when the magnetic pressure,  $\frac{1}{2}\mu_0 H^2$  is significantly larger than any difference in hydrostatic pressure. It can then be linearized when  $M = \chi H$ , under the geometrical conditions  $w \leq \frac{1}{2}h$  (Fig. 1a) and  $d \leq \frac{1}{2}w$ , typical of our devices, where  $w$  is the spacing between the magnets. The linear model gives the minimum equilibrium dimensionless diameter  $d^* = d/w$  as:

$$d^* = \left[ \frac{\pi^2}{N_D \chi (\chi + 1)} \right]^{1/3} \quad (4)$$

**Table 1 | Characteristics of the magnetic fluids**

Magnetic liquid	Susceptibility, $\chi$	Viscosity, $\eta$ (mPa s)*	Appearance
Magoil	$4.7 \times 10^{-4}$	600	Transparent
APG311 ferrofluid	0.151	70	Opaque/black
EMG900 ferrofluid	1.67	60	Opaque/black
MKC ferrofluid	1.88	380	Opaque/black
MD4 ferrofluid	0.315	180	Opaque/black

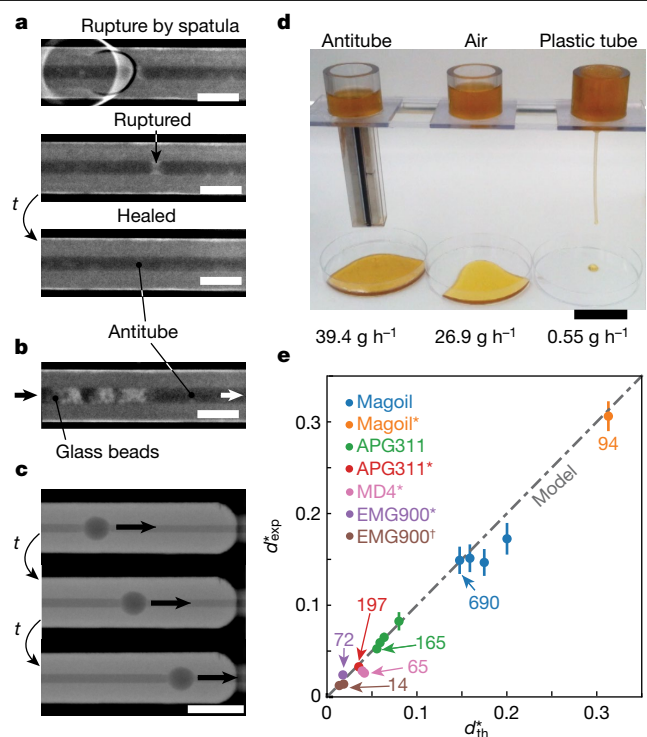
\*At a shear rate of  $100 \text{ s}^{-1}$ 

where  $N_D = \mu_0 M_r^2 w / \sigma$  is the magnetic confinement number expressing the ratio of magnetic to surface energies (see Methods section 'Derivation of linear and saturation models'), and  $M_r$  is the remanent magnetization of the permanent magnets. Note that a 1,000-fold increase in  $\chi$  reduces  $d^*$  by a factor of 100, revealing the importance of the magnetic susceptibility of the confining fluid.

We have studied how the attainable antitube diameters vary with  $w$  for the paramagnetic confinement liquids given in Table 1. There is good agreement between the experimental points for water antitubes and the predictions of equations (3) and (4), using measured susceptibility  $\chi$  and interface energy  $\sigma$  (Extended Data Table 1). The addition of a surfactant to the water phase, to the magnetic fluid phase or to both<sup>20</sup> leads to smaller antitubes since  $\sigma$  is lowered; 1 vol% or 23 mM Span-80, in the ferrofluid, and 1 vol% or 0.58 mM Tween-20 in the aqueous antitube. Although the critical micelle concentrations of Span-80 and Tween-20 in water–hydrocarbon interfaces are 1.2 mM and 0.58 mM, respectively<sup>21</sup>, we did not observe surfactant-induced structure formation. The close agreement is shown in the plot of dimensionless diameters  $d^*$  in Fig. 2e where plots of the experimental values versus predicted diameters collapse onto one curve for all ferrofluids, surface tensions and magnet gaps used. The smallest antitube that could be detected so far has  $d = 14 \pm 2 \mu\text{m}$ , obtained when using a strong ferrofluid, EMG900 (Extended Data Fig. 3g). Our model predicts that antitubes below  $1 \mu\text{m}$  can be stabilized, with a 100- $\mu\text{m}$  magnet spacing and a strong ferrofluid (MKC), but they are currently below the detection limit of our imaging methods.

Magnetic confinement can be used to implement basic microfluidic operations. To make branched antitube devices, we resort to out-of-plane quadrupolar fields, made by waterjet-cutting two stacked  $\text{Nd}_2\text{Fe}_{14}\text{B}$  plates (Fig. 3a). The null-field line follows the channel centre independently of the channel angle with respect to the magnets, which is not the case for in-plane quadrupoles (not shown). Symmetric splitting of the flow was demonstrated in a ferrofluid antitube Y-junction (Supplementary Video 7). Merging of the flow at a Y-junction was visualized for antitubes stabilized by Magoil. Remarkably, merging and rapid mixing occurs immediately after a Y-junction (Fig. 3b), similar to mixing in magnetically stabilized aqueous paramagnetic tubes surrounded by water<sup>16</sup>. This behaviour is in contrast to the laminar flow observed in a 3D-printed microfluidic chip with the same channel size and geometry as the antitube (Fig. 3c).

As our microfluidic circuits are magnetically defined, we can exploit the fact that structured magnetic fields can be modulated by mechanical or electrical means to impose unique and versatile control of fluidic devices. Valves can be constructed by moving one or two longitudinally magnetized bars towards the quadrupole axis. The valving magnets pinch off the antitube by removing the null field at the centre (Fig. 3d, e), thereby imposing a local magnetic pressure,  $P = \mu_0 \mathbf{M} \cdot \mathbf{H}$ , and interrupting the liquid flow (see Supplementary Videos 8–10, which plot the calculated 150-mT isovolume surfaces). The burst pressure of these valves was tested with an Elvex OB1 pressure-driven pump. A single transverse valving magnet could sustain a pressure of 125 mbar, whereas a dual valve (Fig. 3f) withstood 300 mbar. Pumping is an extension of the valving principle; pinch points of magnetic pressure seal off pockets of pumped liquid (for example, water) which are then

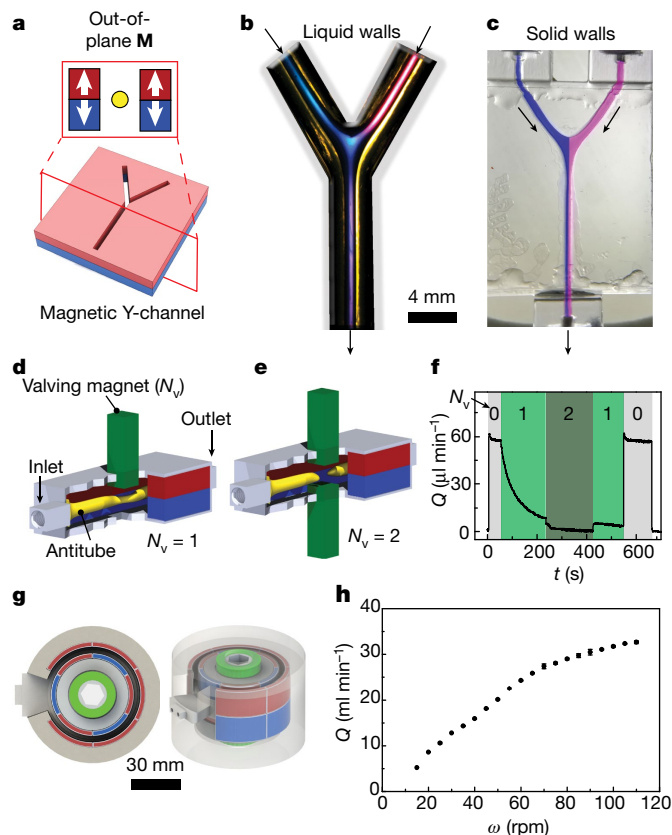


**Fig. 2 | Properties and scaling of water antitubes.** **a**, Top-view X-ray images illustrating the mechanical rupture of a static tube by a spatula, followed by self-healing. The tube returns to equilibrium, without being under flow, within minutes. **b**, Glass beads (0.6 mm) inserted into a 1.5-mm antitube can be expelled with a slight increase in applied pressure of 20 mbar. **c**, A 2-mm-diameter bead, larger than the antitube diameter ( $d = 0.5 \text{ mm}$ ), does not cause clogging. There is only a small decrease in flow rate (from  $30 \mu\text{l min}^{-1}$  to  $23 \mu\text{l min}^{-1}$ ) when the bead is present. In **a–c**, the scale bars are 5 mm. **d**, Comparative flow of honey under gravity (see Supplementary Video 4) through an antitube (left), a normal tube of the same diameter (right) and in free fall (centre; tube open across full diameter). Scale bar, 4 cm. **e**, Plot of experimental dimensionless antitube diameters  $d^* = d/w$  (where  $w$  is the spacing between the magnets) versus theoretical values calculated from equation (4), for a series of magnetic fluids. Numbers indicate the smallest tube diameters in micrometres. \*Water containing 1 vol% Tween-20. †1 vol% Span-80 in the ferrofluid in addition to 1 vol% Tween-20 in the water. For ferrofluid data, each data point represents the average of at least two experiments, for which at least five images were recorded each. For Magoil\*, at least 10 snapshots were taken from an experimental video. Error bars are standard deviations.

displaced by the mechanical actuation of the valving magnets. This can be done by sequential excitation of electro- or electro-permanent magnets, or a rotating array of magnetic spokes (see Extended Data Fig. 6d–f). A more sophisticated pump based on radially magnetized arc-segments fixed onto a rotor and stator, and a 3D-printed fluidic support, is illustrated in Fig. 3g, with magnetic field contours shown in Extended Data Fig. 6j. This ‘Qpump’, using MKC ferrofluid, produces pressures of up to 900 mbar and flow rates of  $32.7 \pm 0.3 \text{ ml min}^{-1}$  (Supplementary Video 12, Fig. 3h).

Because the antitube inside the Qpump has no solid walls, we expected that magnetostatic pumping would be gentler than peristaltic pumping, in which a plastic tube is mechanically squeezed by a roller. It is known that blood pumping results in haemolysis: that is, shear-induced rupture of red blood cells that releases haemoglobin<sup>22–26</sup>. High concentrations of free haemoglobin are cytotoxic and have been associated with clinical complications including an increased incidence of thrombosis, morbidity and mortality<sup>22–26</sup>. We therefore compared the effects of pumping 6-ml samples of human donor blood (collected on hirudin anticoagulant) at  $1.5 \text{ ml min}^{-1}$  for 1 h in a closed



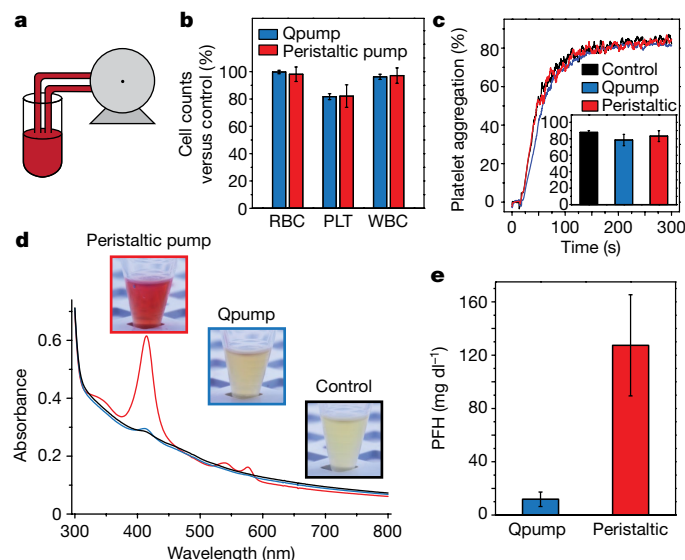


**Fig. 3 | Magnetically implemented mixing, valving and pumping.**

**a**, Out-of-plane magnetization configuration for a Y-junction cut in a double magnet sheet. **b**, Optical image of an aqueous antitube stabilized by  $\text{Tb}^{3+}$ -Magoil. Blue and pink dyes are introduced at the inlets ( $300 \mu\text{l min}^{-1}$ ) and mix immediately upon contact before flowing to the outlet. **c**, Comparison with a normal microfluidic channel, in which no mixing is observed. **d**, **e**, Valving with one magnet ( $N_v = 1$ , Supplementary Video 9; **d**) or two magnets ( $N_v = 2$ , Supplementary Video 10; **e**). **f**, Measured flow rate  $Q$  under an applied pressure of 100 mbar at the exit port controlled by one or two valving magnets. **g**, Top view and isometric view of a magnetostaltic Qpump with rotating magnetic segments. The orientation of magnetization for the arc segments is radially outward (red) or inward (blue). **h**, Flow rate of water versus rotation rate,  $\omega$ , of the inner rotor. Error bars represent the standard deviations over three samples.

loop, using either the Qpump or a peristaltic pump (Fig. 4a). Whereas for water sub-ppm transfer of magnetite nanoparticles was observed, for blood  $285 \pm 161 \text{ ppm}$  ( $\mu\text{g Fe}_3\text{O}_4$  per g blood) was detected (see Methods sections ‘Quantitative analysis of water contamination’ and ‘Blood pumping and analysis’). However, this did not affect haematological parameters such as haematocrit and cell counts of the pumped blood, which were unchanged compared with the peristaltic pump (Fig. 4b and Extended Data Table 2). Moreover, functions of platelets that circulated through the Qpump were unchanged, as they responded normally to agonists such as thrombin-receptor-activating peptides (TRAP) in the widely used light transmission aggregation test (Fig. 4b). Further inspection of platelet morphology by scanning electron microscopy (SEM) confirmed that platelets were intact and normal after pumping (Extended Data Fig. 7a–c). In short, these experiments revealed no adverse effects on whole blood pumped using the Qpump, as compared with a peristaltic pump.

In contrast, a large difference was found in the degree of haemolysis (Fig. 4d, e, Extended Data Fig. 7d). Platelet-rich plasma (PRP) obtained by centrifugation of whole blood is transparent after pumping using the Qpump, but it is bright red after peristaltic pumping (insets,



**Fig. 4 | Blood pumping by a magnetostaltic pump.** Blood was circulated in a closed-loop circuit either by the Qpump (filled with  $7\times$  diluted MKC ferrofluid) or a normal peristaltic pump. **a**, Closed-cycle pumping scheme used.

**b**, Blood-cell counts (RBC, red blood cells; PLT, platelets; WBC, white blood cells) after pumping normalized versus control blood. **c**, The degree of platelet aggregation activated by TRAP tested for PRP from blood after peristaltic pumping and Qpumping, compared with control blood. Light transmission is minimal before aggregation as platelets scatter light, but upon addition of TRAP the platelets aggregate and transmission increases. **d**, Absorption spectrum of PRP from blood after peristaltic pumping and Qpumping, compared with control blood, 25 times diluted with phosphate buffer. **e**, Plasma-free haemoglobin (PFH) concentration estimated from the absorbance of PRP for Qpump and peristaltic pump. Blood was used from three different anonymous donors (see Methods section ‘Blood pumping and analysis’), and bars show data from at least six measurements. Error bars show standard deviations.

Fig. 4d). From the ultraviolet–visible (UV–vis) absorbance of haemoglobin<sup>27</sup> (Fig. 4d), the concentration of plasma-free haemoglobin (PFH) was determined to be  $130 \pm 40 \text{ mg dl}^{-1}$  for the peristaltic pump and  $12 \pm 5 \text{ mg dl}^{-1}$  for the Qpump (Fig. 4e). The degree of haemolysis is  $11 \pm 6$  times lower for blood pumped by the Qpump than for the traditional peristaltic pump. PFH values greater than  $20 \text{ mg dl}^{-1}$  are indicative of high haemolysis<sup>28</sup>, and previous reports on transfusion of packed red blood cells by means of infusion pumps showed PFH values 2–3 orders of magnitude higher<sup>26</sup>. This indicates that magnetic pumping is a promising way to achieve gentle transportation of blood, or possibly other fragile entities such as antibodies or stem cells. Further studies are planned with a mouse model of extracorporeal membrane oxygenation, using a Qpump instead of a peristaltic pump, to investigate survival rates and possible systemic responses upon ferrofluid contact.

Magnetic control of liquid-in-liquid flow opens new possibilities for microfluidics, allowing new channel shapes and low-pressure cargo transport surpassing the current capabilities of standard methods. We have identified the basic physical quantities controlling the size of confined diamagnetic fluid circuitry and have given examples of key microfluidic device elements. The method promises low-shear flow and pumping, which is of growing importance in biotechnology where delicate cells, proteins and antibodies are commonly damaged by traditional pumps<sup>29–33</sup>. If scaled up to  $5\text{--}7 \text{ l min}^{-1}$  and combined with a gradient–field separator<sup>34</sup> to remove traces of transferred magnetite nanoparticles, magnetic blood pumping might be implemented in heart–lung machines during cardiopulmonary bypass surgery, or in devices for extracorporeal membrane oxygenation<sup>22–25</sup>. At a more fundamental level, we envisage miniaturized fluidic circuits without



solid walls that will be scalable down to the submicrometre level. We can then take advantage of the versatility of magnetic control at the nanoscale to open the door to practical low-pressure nanofluidics<sup>35–37</sup>.

## Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-020-2254-4>.

1. Tabeling, P. *Introduction to Microfluidics* (Oxford, 2005).
2. Mukhopadhyay, R. When microfluidic devices go bad. *Anal. Chem.* **77**, 429A–432A (2005).
3. Zhao, B., Moore, J. S. & Beebe, D. J. Surface-directed liquid flow inside microchannels. *Science* **291**, 1023–1026 (2001).
4. Wong, T.-S. et al. Bioinspired self-repairing slippery surfaces with pressure-stable omniphobicity. *Nature* **477**, 443–447 (2011).
5. Wang, W. et al. Multifunctional ferrofluid-infused surfaces with reconfigurable multiscale topography. *Nature* **559**, 77–82 (2018).
6. Leslie, D. C. et al. A bioinspired omniphobic surface coating on medical devices prevents thrombosis and biofouling. *Nat. Biotechnol.* **32**, 1134–1140 (2014).
7. Forth, J. et al. Reconfigurable printed liquids. *Adv. Mater.* **30**, 1707603 (2018).
8. Secchi, E. et al. Massive radius-dependent flow slippage in carbon nanotubes. *Nature* **537**, 210–213 (2016).
9. Banerjee, A., Kreit, E., Liu, Y., Heikenfeld, J. & Papautsky, I. Reconfigurable virtual electrowetting channels. *Lab Chip* **12**, 758 (2012).
10. Choi, K., Ng, A. H. C., Fobel, R. & Wheeler, A. R. Digital microfluidics. *Annu. Rev. Anal. Chem.* **5**, 413–440 (2012).
11. Lee, W. C., Heo, Y. J. & Takeuchi, S. Wall-less liquid pathways formed with three-dimensional microring arrays. *Appl. Phys. Lett.* **101**, 114108 (2012).
12. Walsh, E. J. et al. Microfluidics with fluid walls. *Nat. Commun.* **8**, 816 (2017).
13. Keerthi, A. et al. Ballistic molecular transport through two-dimensional channels. *Nature* **558**, 420–424 (2018).
14. Shang, L., Cheng, Y. & Zhao, Y. Emerging droplet microfluidics. *Chem. Rev.* **117**, 7964–8040 (2017).
15. Zhao, W., Cheng, R., Miller, J. R. & Mao, L. Label-free microfluidic manipulation of particles and cells in magnetic liquids. *Adv. Funct. Mater.* **26**, 3916–3932 (2016).
16. Coey, J. M. D., Aogaki, R., Byrne, F. & Stamenov, P. Magnetic stabilization and vorticity in submillimeter paramagnetic liquid tubes. *Proc. Natl Acad. Sci. USA* **106**, 8811–8817 (2009).
17. Caravan, P., Ellison, J. J., McMurry, T. J. & Lauffer, R. B. Gadolinium(III) chelates as MRI contrast agents: structure, dynamics, and applications. *Chem. Rev.* **99**, 2293–2352 (1999).
18. Ferrand, J., Favreau, L., Joubaud, S. & Freyssingeas, E. Wetting effect on Torricelli's law. *Phys. Rev. Lett.* **117**, 248002 (2016).
19. Rosensweig, R. E. *Ferrohydrodynamics* (Dover, 2014).
20. Posocco, P. et al. Interfacial tension of oil/water emulsions with mixed non-ionic surfactants: comparison between experiments and molecular simulations. *RSC Adv.* **6**, 4723–4729 (2016).
21. Owusu Apenten, R. K. & Zhu, Q.-H. Interfacial parameters for selected Spans and Tweens at the hydrocarbon–water interface. *Food Hydrocoll.* **10**, 27–30 (1996).
22. Byrnes, J. et al. Hemolysis during cardiac extracorporeal membrane oxygenation: a case-control comparison of roller pumps and centrifugal pumps in a pediatric population. *ASAIO J.* **57**, 456–461 (2011).
23. Omar, H. R. et al. Plasma free hemoglobin is an independent predictor of mortality among patients on extracorporeal membrane oxygenation support. *PLoS ONE* **10**, e0124034 (2015).
24. Dalton, H. J. et al. Factors associated with bleeding and thrombosis in children receiving extracorporeal membrane oxygenation. *Am. J. Respir. Crit. Care Med.* **196**, 762–771 (2017).
25. Valladolid, C., Yee, A. & Cruz, M. A. von Willebrand factor, free hemoglobin and thrombosis in ECMO. *Front. Med.* **5**, 228 (2018).
26. Wilson, A. M. M. et al. Hemolysis risk after packed red blood cells transfusion with infusion pumps. *Rev. Lat. Am. Enfermagem* **26**, e3053 (2018).
27. Prael, S. Optical absorption of hemoglobin. *Oregon Medical Laser Center* <https://omlc.org/spectra/hemoglobin/index.html> (1999).
28. Baskin, L., Dias, V., Chin, A., Abdullah, A. & Naugler, C. in *Accurate Results in the Clinical Laboratory* (eds Dasgupta, A. & Sepulveda, J. L.) 19–34 (Elsevier, 2013).
29. Jaouen, P., Vandanjon, L. & Quéméneur, F. The shear stress of microalgal cell suspensions (*Tetraselmis suecica*) in tangential flow filtration systems: the role of pumps. *Bioresour. Technol.* **68**, 149–154 (1999).
30. Kamaraju, H., Wetzel, K. & Kelly, W. J. Modeling shear-induced CHO cell damage in a rotary positive displacement pump. *Biotechnol. Prog.* **26**, 1606–1615 (2010).
31. Vázquez-Rey, M. & Lang, D. A. Aggregates in monoclonal antibody manufacturing processes. *Biotechnol. Bioeng.* **108**, 1494–1508 (2011).
32. Wang, S. et al. Shear contributions to cell culture performance and product recovery in ATF and TFF perfusion systems. *J. Biotechnol.* **246**, 52–60 (2017).
33. Nesta, D. et al. Aggregation from shear stress and surface interaction: molecule-specific or universal phenomenon? *Bioprocess Int.* **30**, 30–39 (2017).
34. Hejazian, M., Li, W. & Nguyen, N.-T. Lab on a chip for continuous-flow magnetic cell separation. *Lab Chip* **15**, 959–970 (2015).
35. Eijkel, J. C. T. & van den Berg, A. Nanofluidics: what is it and what can we expect from it? *Microfluid. Nanofluidics* **1**, 249–267 (2005).
36. Bocquet, L. & Charlaix, E. Nanofluidics, from bulk to interfaces. *Chem. Soc. Rev.* **39**, 1073–1095 (2010).
37. Bocquet, L. & Tabeling, P. Physics and technological aspects of nanofluidics. *Lab Chip* **14**, 3143–3158 (2014).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2020

## Methods

### Synthesis of rare-earth oil

Inspired by the well-known gadolinium complexes of DTPA, used as paramagnetic contrast agents in magnetic resonance imaging, we prepared an amphiphilic complex of DTPA with the paramagnetic rare-earth ion holmium(III), bearing two hydrophobic side chains connected to the DTPA moiety through ester linkage. In particular, by reaction of DTPA dianhydride with the branched alcohol 2-butyl-1-octanol we obtained a tricarboxylic chelating ligand, which on complexation with  $\text{Ho}^{3+}$  under alkaline conditions afforded a stable neutral  $\text{Ho}$ -DTPA complex. The latter, insoluble in water, was mixed with 2-butyl-1-octanol (30 wt% of the latter) affording a water-immiscible homogeneous fluid (Magoil) with positive magnetic susceptibility. The magnetic properties of the neutral  $\text{Ho}$ -DTPA complex (that is, before mixing with the alcohol) were evaluated by NMR measurement, using the Evans method<sup>38</sup>, giving a value  $4.40 \times 10^{-7} \text{ m}^3 \text{ mol}^{-1}$  for the molar susceptibility, which is comparable with that measured for the inorganic salt  $\text{Ho}(\text{AcO})_3$ , as well as of the order of magnitude expected for  $\text{Ho}^{3+}$  (ref. 39). Our paramagnetic oil is a transparent non-ionic liquid, slightly pink-yellowish because of the fluorescence of  $\text{Ho}(\text{III})$ , and it is physico-chemically stable in contact with aqueous solutions for several hours. Over longer periods of contact, hydrolysis occurs. Overall, these characteristics make it different from other magnetic fluids such as ferrofluids (black colloidal suspensions of nanometre-sized magnetic particles) or magnetic room-temperature ionic liquids. The preparation of the oil was easily scaled up to 25 g. We also used different trivalent rare-earth ions, such as erbium and terbium (Fig. 3b), obtaining analogous results.

### Preparation of the $\text{Ho}$ -DTPA-(2-butyl-1-octanol)<sub>2</sub>-based paramagnetic oil

DTPA dianhydride (5 g, 14 mmol) was dispersed in dry dimethylformamide (100 ml) under argon atmosphere, and the suspension obtained was heated to 70–75 °C under stirring, until complete solubilization was observed. After that, 2-butyl-1-octanol (5.5 g, 28 mmol, 2 eq.), dissolved in 10 ml DMF, was added dropwise to the above solution and the resulting homogeneous mixture was left under stirring at 50 °C for 4 h. The completeness of the reaction was monitored by liquid chromatography/mass spectrometry. Afterwards, the solvent was removed (rotary evaporation plus high-vacuum pump) to give **2** as a sticky yellowish solid that was used without further purification (Extended Data Fig. 2a).

To a stirred 0.2 M solution of the ligand **2** in ethanol (EtOH), an aqueous 1 M solution of NaOH (3 eq.) was added dropwise at room temperature, followed by the addition of  $\text{HoCl}_3 \cdot (6\text{H}_2\text{O})$  (1.5 eq. with respect to **2**), dissolved in water to produce a final 0.1 M solution of the complex of **3** in ethanol:water 1:1. The resultant homogeneous mixture was left stirring at room temperature until a viscous pink oil, which responds to magnetic fields, separated out from the water phase (typically 0.5–1 h). After that, the solution was concentrated by rotary evaporation to remove EtOH, and the water phase was extracted with dichloromethane. The organic phase was separated, dried over  $\text{MgSO}_4$  and rotary-evaporated to give the complex **3** as a glassy solid. Finally, the latter was dissolved in the minimum amount of dichloromethane, and the fatty alcohol 2-butyl-1-octanol (30 wt%) was added to the resultant solution. The mixture was then rotary-evaporated under high vacuum to eliminate dichloromethane, resulting in the target homogeneous transparent paramagnetic oil. The absorbance of the paramagnetic oil (Extended Data Fig. 2b) matches the absorption peaks of  $\text{Ho}^{3+}$  (ref. 40).

### Measurement of the magnetic susceptibility of the paramagnetic oil

The magnetic susceptibility of the neutral complex **3** (Extended Data Fig. 2a) was measured by the Evans method, using NMR spectroscopy<sup>38</sup>. This method relies on the fact that the chemical shift of the  $^1\text{H}$  NMR signals of a molecule depend on the bulk susceptibility of the medium.

We used the resonance line of the residual proton of  $\text{CDCl}_3$  (that is, the NMR solvent) as a reference, by comparing its chemical shift in pure  $\text{CDCl}_3$  and in a  $\text{CDCl}_3$  solution containing the complex **3**. From the difference in the chemical shifts we calculated the mass and molar susceptibilities of **3** ( $\chi_{\text{mass}} = 4.88 \times 10^{-7} \text{ m}^3 \text{ kg}^{-1}$ ;  $\chi_{\text{mol}} = 4.40 \times 10^{-7} \text{ m}^3 \text{ mol}^{-1}$ ).

### Magnetic properties of the ferrofluids

Magnetization measurements were performed in a vibrating sample magnetometer with a compact variable permanent magnet source<sup>41</sup>. Samples were mounted in 3D-printed sample holders with cylindrical chambers of diameter 3 mm and length 3.88 mm or 5.88 mm, ensuring a well-defined demagnetization factor and maximal use of the uniform region of the pick-up coils.

The magnetic field dependence of the magnetization was assumed to be due to dispersed non-interacting spherical particles in a non-magnetic medium, and the data were fitted with either a single-component or two-component Langevin function. After correcting for non-zero magnetic field offsets, the applied field  $H$  was corrected for the appropriate demagnetization factor,  $N$  (ref. 42), to get the internal field  $H'$ . Additional magnetization measurements were performed by measuring the inductance of solenoids immersed in the ferrofluids. The solenoid inductance was measured using an LCR bridge meter (Hameg 8118) at 1 kHz. See Supplementary Information section I for Langevin and demagnetization factor equations.

### Remanent magnetization of permanent magnets

The remanent magnetization  $M_r$  of the permanent magnets used was measured by profiling the external field with a Gaussmeter in the  $z$  direction along the central axis and fitting the resultant profile to analytical expressions<sup>43</sup> (see Supplementary Information section II for equations).

### Magnetic field simulations

Assuming magnetically transparent, uniformly magnetized permanent magnets, 3D magnetic fields can be calculated with analytical expressions based on the surface charge model<sup>44,45</sup>. The magnetic field at each point is summed over the contributions of each magnet in an assembly. Full expressions are presented in Supplementary Information section III.

### Surface tension of magnetic liquids

A simple home-built pendent drop set-up was used to measure the surface tension of the magnetic liquids. The solutions were gradually dispensed in a stepwise fashion through a polished 74- $\mu\text{m}$ -diameter glass capillary using a syringe pump into a glass cuvette in a 3D-printed support. The pendent drops were imaged using a Canon EOS 50D and Tamron superzoom 18–270-mm lens in RAW mode. RAW images were converted to monochrome TIFF files, with image processing being carried out in ImageJ<sup>46</sup>, and the surface tension calculated using ‘Pendent\_Drop’, an ImageJ plug-in<sup>47</sup>.

The drops were gradually enlarged, with multiple images taken at each size until the drops broke off, and this was repeated at least six times. The values reported in Extended Data Table 1 are pooled mean and variance, calculated from at least six different droplets, for at least 15 droplet sizes, and three images per size. Overall, between 30 and 100 images of droplets of various sizes were used per surface tension value. Surface tension measurements were cross-checked with a Kruss Drop Shape Analyzer DSA100 and found to be in agreement.

### Method for self-consistent diameter calculation with the full model

For a given set of conditions (that is, permanent magnet remanent magnetization, quadrupole spacing, ferrofluid nanoparticle diameter and volume fraction, and surface tension),  $d$  can be calculated in a self-consistent manner by calculating the magnetic field and magnetization at a position  $d/2$  with the scheme in ‘Remanent magnetization

of permanent magnets', next generating a new  $d$  using equation (3), using this as a new input  $d$ , and looping repeatedly until the difference between input and output is negligible.

## Side-view optical imaging and pressure/flow measurements

Images were acquired with either a Leica MZ16 stereo microscope with a Leica IC80HD digital camera, or a Nikon SMZ745T stereo microscope and a SONY colour CCD camera (1/1.8", 20 images per second,  $1,600 \times 1,200$  pixels). An Elveflow OB1 pressure-driven pump and Elveflow FS4 flow sensors were used to set the applied pressure and measure the resulting flow rates in and out of the antitubes. All image processing and analysis was carried out in ImageJ<sup>46</sup>. Image thresholding was performed using Otsu's clustering method, with the resulting image converted to a binary mask. Particle analysis and circle fitting was performed on each image, resulting in cross-sectional area, and roundness/circularity. Examples of cross-sections are shown in Extended Data Fig. 4d.

## Optical transmission imaging

A Zeiss Axiovert 200M inverted microscope with a  $5\times$  NA = 0.15 objective and Andor Zyla 4.2 sCMOS detector were used to image through 3D-printed microfluidic devices. These devices consist of a main channel 200  $\mu\text{m}$  high and 1 mm wide, a right-angled Y-junction for two water inlets, a ferrofluid inlet which joins the channel at its centre, a single outlet, and space to place four magnets to generate a quadrupolar field centred in the main channel. The device is sealed with a thin glass coverslip.

To stabilize antitubes, two methods were used. (i) The channel was prefilled with water, and the ferrofluid was injected, gradually displacing the water inside the channel; or (ii) the channel was prefilled with ferrofluid, with water injected to displace ferrofluid.

Tube diameter was estimated with ImageJ. Plot profiles were generated by taking column averages of images (Extended Data Fig. 3f, g), followed by subtraction of the background, leading to a sharp peak due to the lower absorption of water. This was fitted with a Gaussian peak function, whose full-width half-maximum (FWHM) corresponded to tube diameters measured from the optical side-view technique.

## Top-view optical imaging through Magoil

A Nikon SMZ745T stereo microscope and a SONY colour CCD camera (1/1.8", 20 images per second,  $1,600 \times 1,200$  pixels) were used to image water tubes inside Magoil. Similar to the X-ray and optical transmission images, the optical images were rotated and inverted (Extended Data Fig. 8), and profiles of column averages were extracted for data fitting. The background was subtracted, and the resulting peaks fitted with a Gaussian function.

## X-ray transmission imaging

A MyRay dental X-ray imaging system was adapted for imaging through ferrofluids. It consisted of an X-ray source (MyRay RX-DC eXtend) and a detector (MyRay Zen X T1) purchased from Castebl (Waver, Belgium), and confined inside a box (constructed in-house) covered by lead plates. In general, X-ray emission at 65 kV and 6 mA with exposure time of 0.1 s was used to obtain a good-contrast image of water antitubes confined in ferrofluids.

To measure the tube diameter, first a background measurement was taken with a microfluidic device (Extended Data Fig. 3a, b) fully filled with ferrofluid. The image processing loop (also in ImageJ) consisted of taking the average of 10 images, rotated to have the tube fully vertical, with the intensity values inverted (Extended Data Fig. 3c). Next a plot profile was generated by taking column averages (Extended Data Fig. 3d, red line). Then a freshly cleaned device was filled with water, and ferrofluid was added progressively in 50  $\mu\text{l}$  steps at a flow rate of 100  $\mu\text{l min}^{-1}$ , with 10 images taken at each step. The same imaging process was carried out as with the background, followed by subtraction

of the background plot profile, leading to a sharp peak due to the lower absorption of water (Extended Data Fig. 3e). This was fitted with a Gaussian peak function, whose FWHM corresponded to tube diameters measured from the optical side-view technique. The procedure of adding ferrofluid was continued until a continuous tube was no longer observed. For improved contrast sensitivity, additional transmission images were taken with a Hamamatsu C7942CK-12 flat panel sensor in place of the MyRay Zen X T1 (Extended Data Fig. 1a, b).

## Synchrotron X-ray tomography

Phase-contrast synchrotron X-ray tomography was carried out at the X02DA TOMCAT beamline of the Swiss Light Source (SLS) at the Paul Scherrer Institute. The X-ray beam, produced by a 2.9-T bending magnet on a 2.4-GeV storage ring (with ring current  $I = 400$  mA), was monochromated to 30 keV. We used a high-speed CMOS (complementary metal-oxide-semiconductor) detector (PCO.Edge 5.5) coupled to a KinoOptik  $1\times$  visible-light objective with a 300- $\mu\text{m}$ -thick LuAG:Ce scintillator, for a field of view 16.6 mm  $\times$  14.0 mm and effective pixel size 6.5  $\mu\text{m}$ . The sample-to-detector distance was 327 mm. Raw images were acquired with 50-ms exposure times across 1,500 tomographic projections. Twenty dark and 100 flat-field images were collected for each scan.

The samples consisted of 3D-printed supports with circular cross-section and a 1.5-mm-diameter channel. There is a groove on the top and bottom of the holder to accommodate two N42 grade Nd<sub>2</sub>Fe<sub>14</sub>B magnets (10  $\times$  10  $\times$  10 mm) on each side at a distance of 3 mm. The 3D print is designed so that the centre of the null field is close to the centre of the fluidic channel. The channels were prefilled with water containing 1 vol% Tween-20 surfactant and imaged after sequential injection of known volumes of a ferrofluid with a syringe pump.

Before further analysis, each projection was corrected with the respective dark and flat-field image. Computed tomography reconstruction was performed with the 'gridrec' algorithm<sup>48</sup>, enabling fast reconstructions of large datasets. Tomographic slices were rendered using ImageJ 1.52i, and the volumetric cross-sections were rendered using Paraview 5.6.0.

## Antitube durability

The durability of the antitubes was tested by varying the flow rate of water with a syringe pump to determine the threshold value at which ferrofluid is sheared out. For this test, we used a fluidic chip that consists of four magnets (each 9  $\times$  9  $\times$  50 mm) with gaps of 9 mm between magnets. No major leakage of ferrofluid was observed throughout the measurement, and the threshold value was reported for the flow rate when iridescence due to a thin film of oil on the water surface was observed. The observed data (Extended Data Fig. 4a, b) are valid only for this specific magnet arrangement and fluidic circuit, as the magnetic field distribution varies when the size and arrangement of magnets are changed, and outlets perpendicular to the main channel inhibit further leakage. In general, the threshold value increases when more viscous or magnetically stronger ferrofluids are used for antitube formation and decreases as the antitube diameter decreases.

## Quantitative analysis of water contamination by magnetic media

The amount of magnetic media leaked into water was quantified by atomic absorption spectroscopy (AAS; Agilent Technologies 200 Series AA). The emission wavelengths of iron (372 nm) and holmium (559 nm) were used for ferrofluids and Magoil, respectively. The standard samples for iron were prepared by dissolving iron metal in hydrochloric acid and diluted by distilled water to obtain a series of concentrations to construct a calibration curve. Each point is from five consecutive measurements. We mixed 1 ml of ferrofluid (MKC and MD4) and 25 ml of distilled water in a separatory funnel and left them overnight. We then transferred 20 ml of the distilled water from the funnel to a vial and removed water by using

a rotary evaporator. Hydrochloric acid (1 ml) was added to the vial to dissolve any remaining iron. The iron solution was diluted to make 50 ml in total and used for AAS. Similarly, we mixed 1 ml of Magoil with 25 ml of distilled water. Five consecutive measurements were performed for each sample to build the statistics. Emission intensities of  $0.049 \pm 0.002$  and  $0.050 \pm 0.004$  were observed from the water samples in contact with MKC and MD4, respectively. Using the calibration curve and error propagation analysis, the concentration of the iron from water in contact with MKC and MD4 was determined as  $0.78 \pm 0.05$  ppm and  $0.79 \pm 0.08$  ppm.

Additional checks were made by flowing water through antitubes of three different diameters, 3.5 mm, 1.5 mm and 0.8 mm. The cavity containing both ferrofluid and water was 6 mm in diameter. We used DI water as transported fluid. The surrounding ferrofluid was APG1141 with viscosity 5 Pa s. Three flow rates were used, corresponding to 10%, 50% and 75–80% of the threshold flow rate for each diameter ( $Q_{\text{thres}} = 80 \text{ ml min}^{-1}$ ,  $60 \text{ ml min}^{-1}$  and  $30 \text{ ml min}^{-1}$ , respectively; see Extended Data Fig. 4a, b). The set-up consists of a pressure-driven pump (Elvesys OB1) connected to a water reservoir which is then connected to the antitube housed in a 3D-printed support. At the outlet, the flow is split into two streams. One passes directly to a tube, and the other flows adjacent to a magnet ( $6 \times 6 \times 50 \text{ mm}$ ) before entering a second tube. This additional magnet was placed with its long axis parallel to the channel flow and is there to capture any ferrofluid that may inadvertently shear out under flow. The water was flowed continuously for 1.5 h, and 45-ml samples were collected in each Falcon tube at the beginning, middle and end of the experiment. All solutions were evaporated, followed by the addition of 1 ml concentrated HCl, which was allowed to react for 1 h and diluted with 45 ml of ultrapure  $\text{H}_2\text{O}$ . The entire procedure was repeated to produce a total of six samples per experimental condition, with the data reported in Extended Data Fig. 4f being the average values, and error bars being the standard deviations. Here we found that all Fe concentrations detected were below 0.3 ppm, with most of them within one standard deviation for a series of blank measurements.

Magoil hydrolyses in a few hours, and therefore 20 ml of water was transferred to a vial after one hour and diluted to make a total 50 ml of the sample for AAS. Holmium(III) chloride was dissolved in distilled water and diluted to obtain a calibration curve. The emission intensity of  $0.058 \pm 0.001$  was observed from the water sample that was in contact with Magoil, which corresponds to  $2.6 \pm 0.1$  ppm based on the calibration curve.

### Construction of Qpump

The Qpump housing was 3D-printed using F170 (Stratasys) with ABS (acrylonitrile butadiene styrene)-M30.  $\text{Nd}_2\text{Fe}_{14}\text{B}$  arc magnets were purchased from NEOTEXX (TR-036.5-33-20-NN and -SN, TR-028.5-25-20-NN and -SN). A high-power gear motor (RobotShop, Devantech 24V) was used to rotate the inner rotor of the Qpump. The fluidic chips were 3D-printed using a Form 2 print (Formlabs) with a clear resin (FLG-PCL02). The chips were then connected to Tygon tubing (Saint-Gobain, ND 100-65, 1/16" inner diameter (ID)  $\times$  1/8" outer diameter (OD)).

### Blood pumping and analysis

Human blood (50 ml) from a unique donor at l'Établissement français du sang (EFS) Strasbourg was used for each set of experiments. The blood was collected on anticoagulant with hirudin ( $100 \text{ units ml}^{-1}$ ). Six 6-ml tubes were prepared for each experiment: one as a negative control (not used in a pump), three for use with a Qpump, and two for a peristaltic pump. To build statistics, blood from three different donors was used to construct the data reported here. The rotation rate of motors was adjusted to obtain the flow rate of each pump at  $\sim 1.5 \text{ ml min}^{-1}$ . Tygon tube (Saint-Gobain, Tygon 3350, 1/16" ID  $\times$  3/16" OD) with stoppers was used with the peristaltic pump (Fischer Scientific, CTP300).

After an hour of blood pumping in a closed circuit (Fig. 4a), the whole blood from each tube was analysed with a haematology analyser (Sysmex Europe, XN-1000). Then the whole blood was centrifuged (SORVALL RC3BP, 10 min at 250 g, room temperature) to obtain PRP

for UV-vis spectroscopy (Agilent, Cary 8454) and light transmission aggregometry (APACT 4004). For UV-vis spectroscopy, PRP was 25 times diluted with phosphate buffer. Light transmission was recorded during 5 min while platelets aggregated after the addition of  $30 \mu\text{l}$  of agonist (TRAP  $10 \mu\text{M}$ ) into  $270 \mu\text{l}$  of PRP.

To perform electron microscopy imaging of platelets, samples were prepared as follows: first,  $100 \mu\text{l}$  of PRP was added to  $400 \mu\text{l}$  glutaraldehyde 2.5% (Euromedex, 16210) in cacodylate (CACO) buffer in a 15-ml tube and incubated overnight at  $4^\circ\text{C}$ . Then the solution was centrifuged for 8 min at  $2,000g$  to obtain pellet platelets, which were then resuspended with 2 ml CACO buffer for washing. This solution was centrifuged again for 8 min at  $2,000g$ , and the collected platelets were resuspended with  $250 \mu\text{l}$  of CACO buffer. Finally, the solution was diluted 10 times, and  $100 \mu\text{l}$  of diluted solution was deposited on a poly-L-lysine (Sigma Aldrich, P8920) coverslip (Euromedex, 72226-01) to sediment the platelets at room temperature. After 30 min, the excess solution was removed, and the substrate was washed twice with CACO buffer for 5 min. Dehydration of the substrate was done with ethanol (LPCR (Les Produits Chimiques de la Robertsau) 20821296): 70% ( $3 \times 5 \text{ min}$ ), 85% (5 min), 95% (5 min) and 100% ( $2 \times 30 \text{ min}$ ). The substrate was then desiccated with hexamethyldisilazane (LPCR, 112186.0100): 1/4 HMDS + 3/4 ethanol 100% (5 min), 1/2 HMDS + 1/2 ethanol 100% (5 min), 3/4 HMDS + 1/4 ethanol 100% (5 min) and pure HMDS ( $2 \times 5 \text{ min}$ ). The coverslip was fixed on an aluminium support mount (Euromedex, 75230) with conductive carbon cement (Euromedex, 12664). The sample was metallized (coating by  $\sim 10\text{-nm}$  layer of platinum and palladium) by a Cressington 208 HR sputter coater and observed with Phenom Pro Desktop SEM for high-quality imaging.

Ferrofluid contamination in the pumped blood was determined using the protocol in 'Quantitative analysis of water contamination by magnetic media', along with a complementary method of measuring the magnetic moment of the whole blood samples in a SQUID magnetometer (Quantum Design MPMS3 SQUID-VSM). First, the ferrofluid MKC was dried by rotary evaporation to measure the magnetic response without the hydrocarbon carrier fluid. Similar to the Langevin function fitting (see 'Magnetic properties of the ferrofluids'), the concentration of superparamagnetic magnetite nanoparticles in the powder was found to be 76%.

Next, eight whole-blood samples, four pumped through two different Qpumps, two pumped through a peristaltic pump, and two controls (not pumped at all), were lyophilized before mounting in gelatine capsules. The control and peristaltic pumped samples showed only a diamagnetic susceptibility at room temperature, as expected for oxyhaemoglobin, with an average susceptibility of  $-2.3 \pm 0.3 \text{ A m}^2 \text{ kg}^{-1} \text{ T}^{-1}$  (after correction for sample holder contributions). In these blood samples, deoxyhaemoglobin is visible only at low temperatures, where a Brillouin function fit reveals an average of four unpaired spins, as expected. In contrast, the four magnetically pumped samples returned an average superparamagnetic moment of  $26.2 \pm 14.8 \times 10^{-3} \text{ A m}^2 \text{ kg}^{-1}$ , which is the equivalent of  $285 \pm 161$  ppm of  $\text{Fe}_3\text{O}_4$  per gram of blood.

### Derivation of equations (1) and (2)

We use the steady-state Stokes equations in radial coordinate for ferrofluid as well as honey. The fluid landscape is divided into three regions, namely the honey flow (region I), ferrofluid with shear flow (region II) and reverse flow of ferrofluid (region III) (see Extended Data Fig. 5e). The solid walls are given a zero-velocity boundary condition, and the interface between the liquids has continuous velocity and shear stress. As there is no net flow of ferrofluid, the flow rate of ferrofluid in region III is equal and opposite to that in region II. This can be seen in the normalized velocity profile in Extended Data Fig. 4e which was also confirmed by numerically solving the Navier–Stokes equation in Extended Data Fig. 5a–d. Finally, solving the Stokes equations with the respective boundary conditions, the effective slip length is defined in terms of the flow rate (for a fixed pressure gradient) in the system. The full derivation is presented in Supplementary Information section III.



## Derivation of equilibrium diameter equation (3)

The approach starts with the magnetically augmented Bernoulli equation, choosing appropriate boundary conditions at two points (see Extended Data Fig. 5f), one at the centre of the antitube in the null field region,  $P_1$ , and the other at the boundary between the inner solution and the magnetic liquid,  $P_2$ , and solving the Bernoulli equation to arrive at an expression due to the balance of energies at the interface. A derivation is provided in Supplementary Information section IV.

## Derivation of linear and saturation models for equilibrium tube diameter

Equation (3) is simplified to equation (4) by linearizing the magnetic field dependence inside the quadrupole, assuming a linear magnetic susceptibility for the ferrofluid, and inserting both results into equation (3). A full derivation is described in Supplementary Information section V.

## Computational fluid dynamics

Finite element methods using ANSYS 18 were used to simulate fluid dynamics of a honey antitube inside EMG900 ferrofluid. A full 3D simulation was performed, with the following boundary conditions applied: at the inlet we define the flow rate ( $Q = 175 \mu\text{l min}^{-1}$ ), and the outlet is set to atmospheric pressure. The interface between liquids is defined as two walls with constant magnitude of shear stress ( $\tau$ ). We vary the shear stress until the velocity at the interface becomes continuous. Once the shear stress and velocity become continuous, the solution is accepted. Here  $\tau = 0.45 \text{ Pa}$  for the semi-infinite case and  $\tau = 0.47 \text{ Pa}$  for the case in which we include the effect of curved inlet and outlet. The reverse flow as predicted by the analytical model is clearly seen in the simulations. The antitube diameter is 1.2 mm and the cavity size is 4.4 mm. The fluids used are honey with viscosity 10 Pa s and EMG900 (ferrofluid) with viscosity 0.06 Pa s. To simplify the simulations, we use free slip conditions at the curved inlet and outlet region. This is because the change in the velocity in the curved inlet and outlet region is dominated solely by the change in diameter.

When comparing an infinite channel (Extended Data Fig. 5a, b) to a finite channel (Extended Data Fig. 5c, d), we observe the same counterflow of ferrofluid. For the finite channel, the velocity away from either the inlet or outlet has a flat, plug-like profile (Extended Data Fig. 5d) identical to the infinite channel (Extended Data Fig. 5b), but at the outlet there is a large drop in the flow velocity as the volumetric flow rate is fixed and the channel widens. Furthermore, most of the pressure drop occurs at both the inlet and outlet where the channel diameter changes (Extended Data Fig. 5c). Additionally, we observe the same pressure drop inside the finite antitube away from the inlet and outlet, as found for the infinite antitube (Extended Data Fig. 5a).

## Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

## Data availability

Source data for Figs. 2, 3 and 4 (and Extended Data figures containing data graphs) are provided with the paper. Any other data that support

the findings of this study are available on the Zenodo data repository, <https://doi.org/10.5281/zenodo.3603029>

## Code availability

The Python code for calculating magnetic fields is available on the Zenodo data repository, <https://doi.org/10.5281/zenodo.3603029>

38. Evans, D. F. The determination of the paramagnetic susceptibility of substances in solution by nuclear magnetic resonance. *J. Chem. Soc.* **1959**, 2003–2005 (1959).
39. Coey, J. M. D. *Magnetism and Magnetic Materials* (Cambridge Univ. Press, 2010).
40. Sastri, V. R., Perumareddi, J. R., Rao, V. R., Rayudu, G. V. S. & Bünzli, J.-C. G. *Modern Aspects of Rare Earths and their Complexes* (Elsevier, 2003).
41. Cugato, O., Byrne, R., McCaulay, J. & Coey, J. M. D. A compact vibrating-sample magnetometer with variable permanent magnet flux source. *Rev. Sci. Instrum.* **65**, 3570–3573 (1994).
42. Wysin, G. M. *Demagnetization Fields* (Kansas State Univ., 2012); <https://www.phys.k-state.edu/personal/wysin/notes/demag.pdf>.
43. Furlani, E. P. *Permanent Magnet and Electromechanical Devices* (Academic, 2001).
44. Yang, Z. J., Johansen, T. H., Bratsberg, H., Helgesen, G. & Skjeltorp, A. T. Potential and force between a magnet and a bulk  $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$  superconductor studied by a mechanical pendulum. *Supercond. Sci. Technol.* **3**, 591 (1990).
45. Camacho, J. M. & Sosa, V. Alternative method to calculate the magnetic field of permanent magnets with azimuthal symmetry. *Rev. Mex. Fis. E* **59**, 8–17 (2013).
46. Schindelin, J. et al. Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).
47. Daerr, A. & Mogne, A. Pendent\_Drop: An ImageJ plugin to measure the surface tension from an image of a pendent drop. *J. Open Res. Softw.* **4**, e3 (2016).
48. Marone, F. & Stamparoni, M. Regriding reconstruction algorithm for real-time tomographic imaging. *J. Synchrotron Radiat.* **19**, 1029–1037 (2012).

**Acknowledgements** We acknowledge the support of the University of Strasbourg Institute for Advanced Studies (USIAS) Fellowship, the ‘Chaire Gutenberg’ of the Région Alsace (J.M.D.C.), the French National Research Agency (ANR) through the Programme d’Investissement d’Avenir under contract ANR-11-LABX-0058\_NIE within the Investissement d’Avenir programme ANR-10-IDEX-0002-02, and SATT Conectus funding. This project has received funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement no. 766007. We acknowledge the Paul Scherrer Institut for provision of synchrotron radiation beamtime at beamline TOMCAT of the SLS. We thank H. Boping of San Huan Corporation for giving us thin magnetic bilayer sheets. We thank F. Chevrier for technical support, and the staff of the STnano nanofabrication facility for help in sample fabrication. We thank N. Matoussevitch for the synthesis of ferrofluids. We thank F. Sacarelli and G. Formon for additional AAS measurements, A. Cebers of the University of Latvia Riga for the use of ANSYS 18, and S. Potier for advice on the project.

**Author contributions** J.M.D.C., B.D. and T.M.H. conceived and initiated the project. T.A. and P.D. performed most of the experiments and modelling. T.A., J.M.D.C., B.D., P.D. and T.M.H. designed the microfluidics set-ups. A.A.D. performed slip length experiments and modelling. P.D. carried out the magnetic modelling. A.S. developed and synthesized Magoil. L.G. characterized liquid and pumping properties. A.B. led the X-ray tomography experiments. T.A., C.B., T.M.H. and P.H.M. designed and conducted the blood experiments. T.A., J.M.D.C., B.D., P.D. and T.M.H. drafted the manuscript. All authors discussed and contributed to the paper in its final form.

**Competing interests** T.M.H. holds shares in Qfluidics, a company devoted to the commercialization of the liquid tube technology presented in this work. P.D., B.D., J.M.D.C. and T.M.H. are co-inventors on patents protecting the technology (WO2018134360A1, pending) for which all parts of the manuscripts are covered.

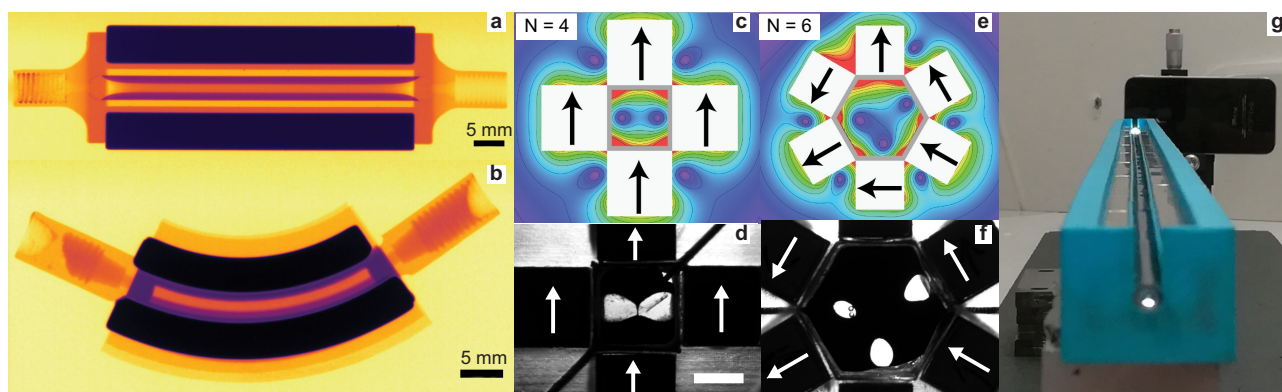
## Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41586-020-2254-4>.

**Correspondence and requests for materials** should be addressed to T.M.H.

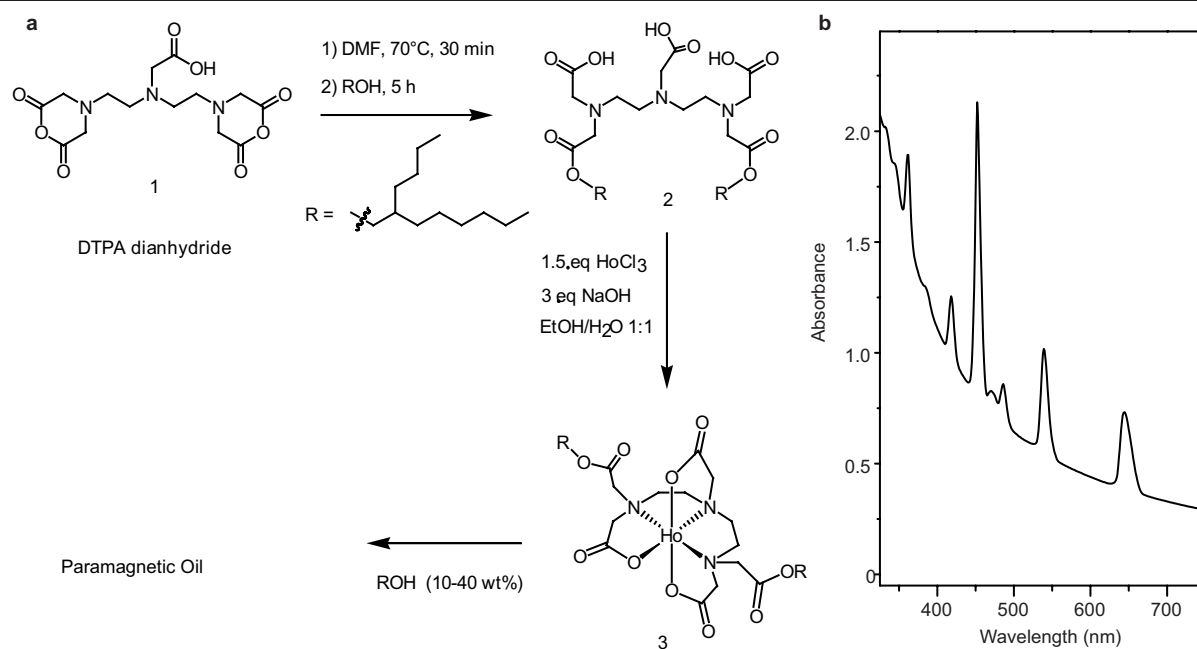
**Peer review information** Nature thanks Emmanuel Delamarque and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>.

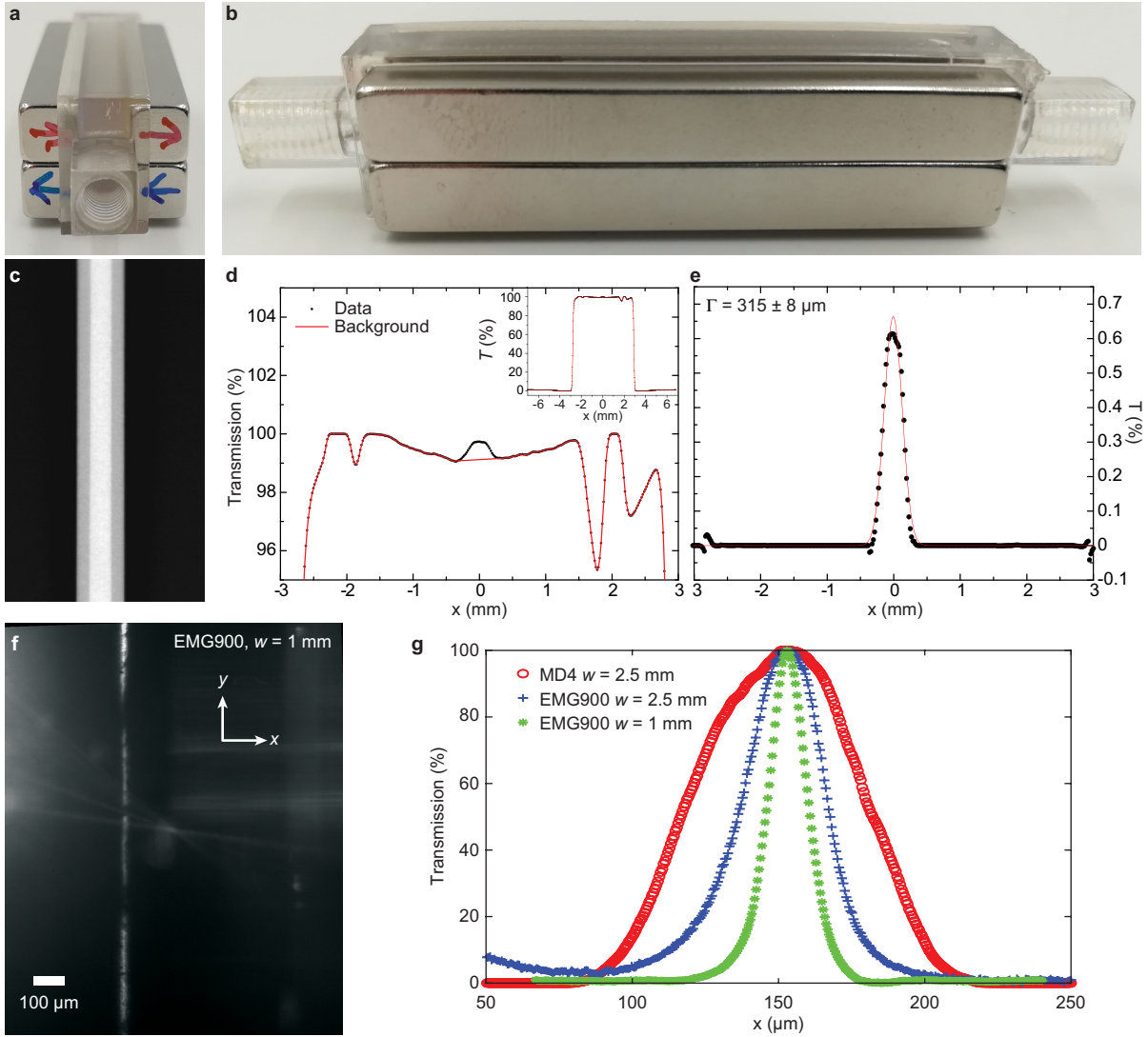


**Extended Data Fig. 1 | Adaptability of antitube topologies to arbitrary magnetic fields.** **a, b**, X-ray transmission images of an antitube of water inside a ferrofluid for cuboid magnets ( $6 \times 6 \times 50$  mm) with 6-mm gap (**a**); and matching arc magnets of height 20 mm, 3.5-mm gap, and inner and outer diameter pairs of: ID 25 mm, OD 28.5 mm; and ID 33 mm, OD 36.5 mm (**b**). **c–f**, Antitube cross-sections using non-quadrupolar fields: the magnetic field

calculation (**c**) and experimental cross-section (**d**) for a four-magnet arrangement; and the magnetic field calculation (**e**) and experimental cross-section (**f**) for a six-magnet arrangement. Scale bar, 3 mm. See Supplementary Videos 5, 6. **g**, Side view of a 1-m-long water antitube ( $d = 2$  mm), which allows light to pass throughout, showing the continuous water phase.



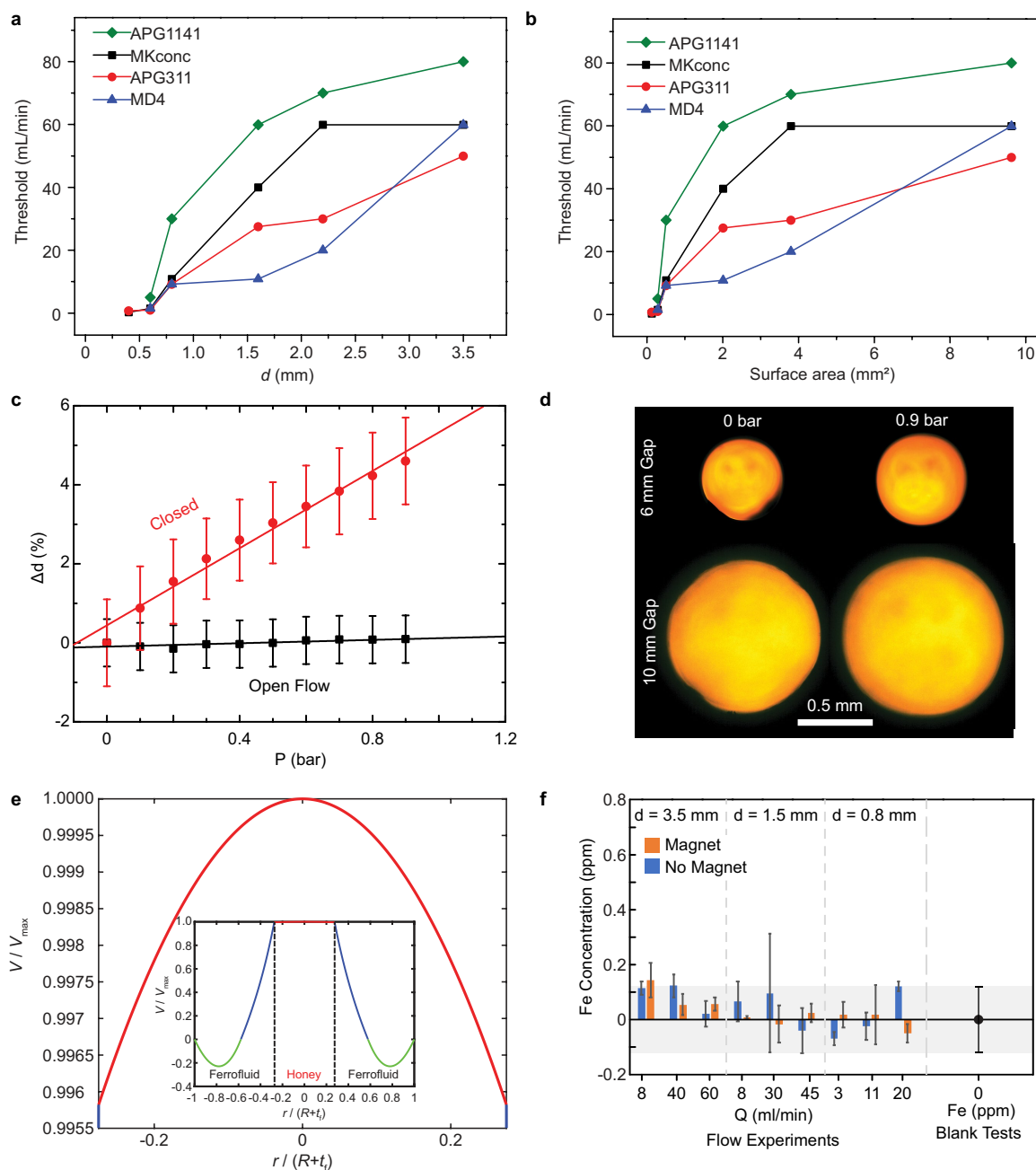
**Extended Data Fig. 2 | Magoil synthesis and optical absorption properties.** **a**, Magoil synthesis reaction scheme. DMF, dimethylformamide; EtOH, ethanol. **b**, UV-vis absorption spectrum of Ho<sup>3+</sup>-based paramagnetic Magoil.



**Extended Data Fig. 3 | Process steps involved in X-ray and optical imaging of antitubes in ferrofluids.** **a, b**, Typical quadrupole assemblies used for X-ray measurements. **c**, Inverted transmission X-ray image. **d**, Transmission averaged along the channel. **e**, Background-corrected transmission through the water

antitube fitted with a Gaussian peak function. **f**, Optical image of a sub-100- $\mu\text{m}$  water antitube in an EMG900 ferrofluid (double surfactant). **g**, Intensity profile across the microfluidic channel in the vicinity of the water antitube. The profile is column-averaged along the length of microfluidic channel.

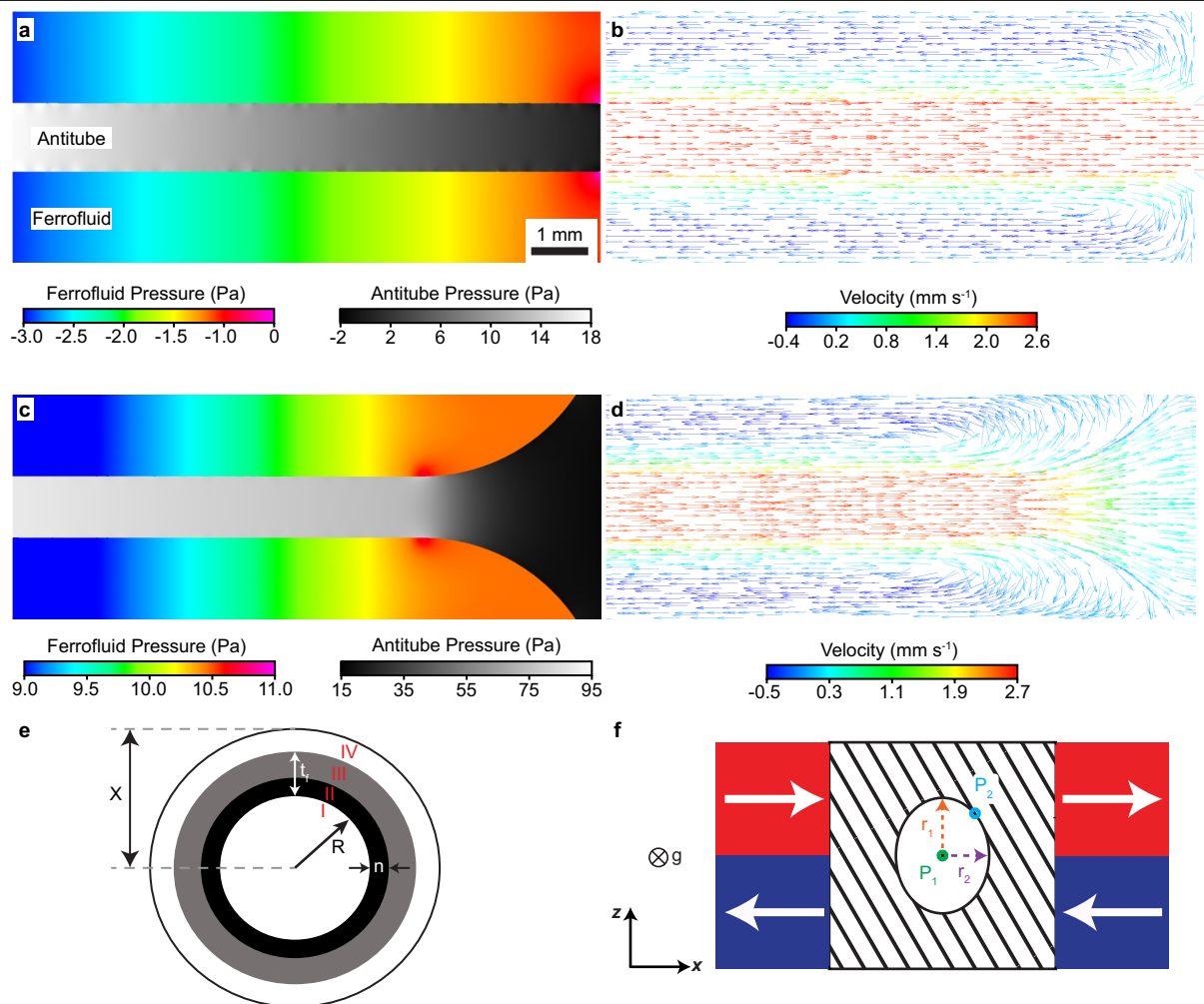




**Extended Data Fig. 4 | Antitube stability and shear resistance under flow.**

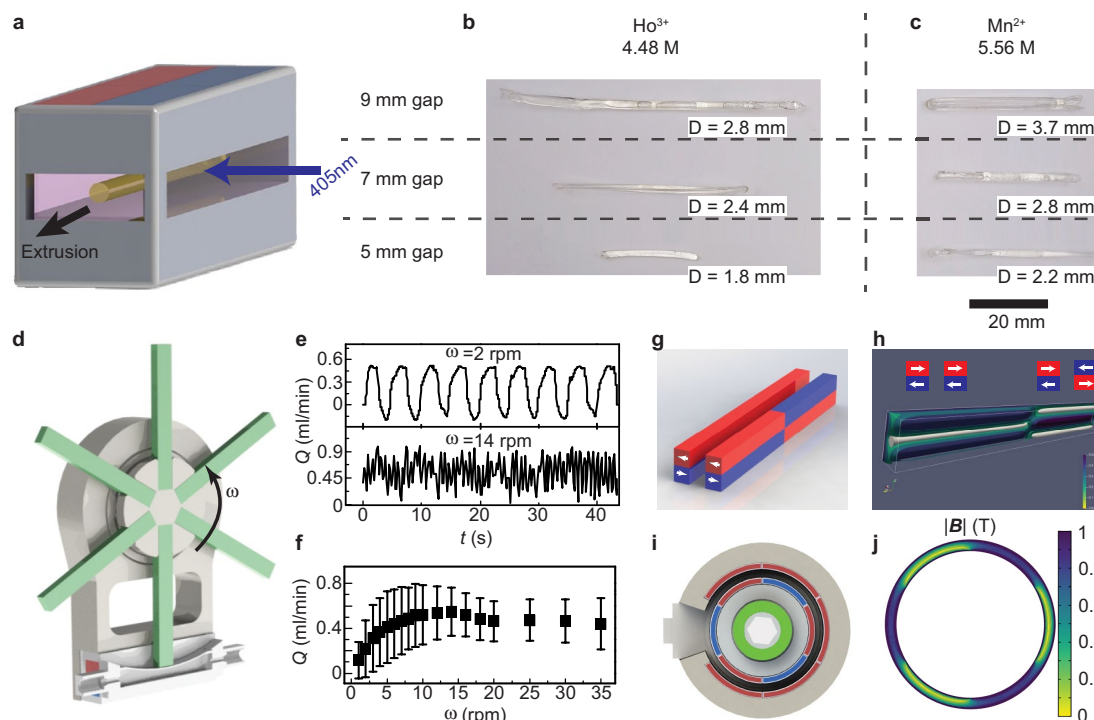
**a, b**, The threshold flow rate for antitube diameters (**a**) and areas (**b**). **c**, Relative dilation of an antitube in APG311 ferrofluid with a quadrupolar gap,  $w$ , of 10 mm under flow and with the outlet closed. **d**, Side view through the antitubes under static conditions with no flow (outlet closed) for two magnet gaps at equilibrium (0 bar) and under 0.9-bar pressure. **e**, Normalized velocity profile

inside honey using equations (1) and (2), corresponding to data in Fig. 2d. Inset shows velocity profile in both honey and ferrofluid. **f**, Concentration of Fe found in water after pumping through three different antitube diameters for three flow rates with and without an extra magnet on the outlet flow. Blank tests for pure water give the grey background threshold detection. Values are averages of six samples, error bars are standard deviations.



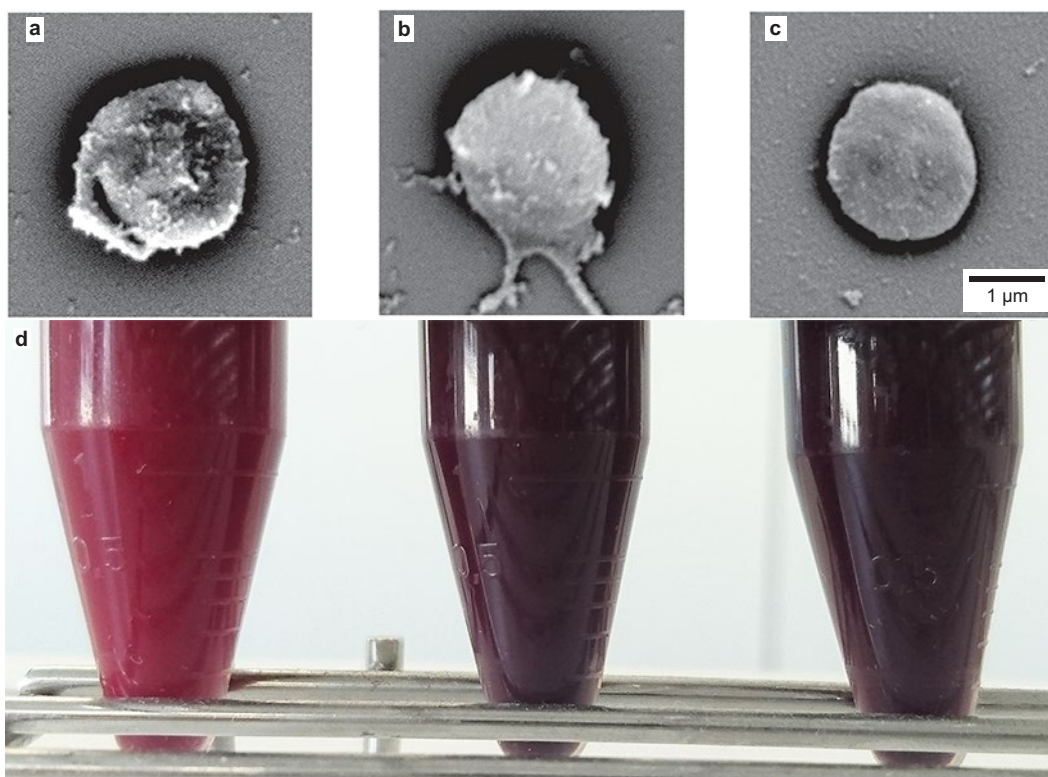
**Extended Data Fig. 5 | Flows in antitubes and surrounding ferrofluid using computational fluid dynamics.** **a–d,** Contour plots from numerical simulations of a honey antitube in EMG900 ferrofluid under a flow rate of  $175 \mu\text{l min}^{-1}$  for two cases: first, the semi-infinite case with no inlet effects (**a, b**); second, the finite case including inlet effects with the ferrofluid contours matching those found by experiment (**c, d**) (compare with Extended Data Fig. 1a). The plots are (**a, c**) isometric pressure contours for at the outlets; note the different colourmaps for the pressure inside the ferrofluid versus inside the antitube; (**b, d**) velocity vector field at the outlets. **e, f,** Geometries used in the derivation of equations (1)–(4). **e,** Cross-section of liquid tube system considered in derivation of equilibrium diameter equation (3) (see Methods),

with four different flow regions consisting of: I, honey; II, a parallel flow of ferrofluid; III, a counterflow of ferrofluid; and IV, a fictitious region to define the radial distance  $x$  at which the flow velocity becomes zero. Thus, the slip length for a flow of honey is  $b = x - R$ , where  $t_f$  is the thickness of the ferrofluid,  $n$  is the thickness of region with shear flow and  $R$  is the radius of the honey tube. **f,** 2D geometry of four bar magnets in a quadrupolar configuration considered in the derivation of linear and saturation models for the equilibrium tube diameter (see Methods and Supplementary Information). The hatched region denotes the ferrofluid, and the white region in the centre is the contained liquid tube.



**Extended Data Fig. 6 | Demonstration of photo-polymerization and pumping functionality in antitubes.** **a**, A scheme of the fluidic chip used for the extrusion of a photopolymer resin and its photopolymerization by 405-nm laser light during extrusion. **b**, **c**, Photos of polymerized tubes extruded in **(b)** an aqueous  $\text{HoCl}_3$  solution, and **(c)** an aqueous  $\text{MnCl}_2$  solution. The diameter of the tubes decreased as the magnet gap decreased, and those photopolymerized in  $\text{Ho}^{3+}$  were smaller, as  $\text{Ho}^{3+}$  has higher magnetic susceptibility than  $\text{Mn}^{2+}$ . **d**, Isometric-view of a six-spoke magnetostatic pump (see Supplementary Video 11). **e**, Slow rotation (2 rpm) leads to pulsed flow, whereas fast rotation (14 rpm) produces a smoother flow. **f**, The average flow rate and standard deviation versus rotation rate  $\omega$ . **g**, The magnetic

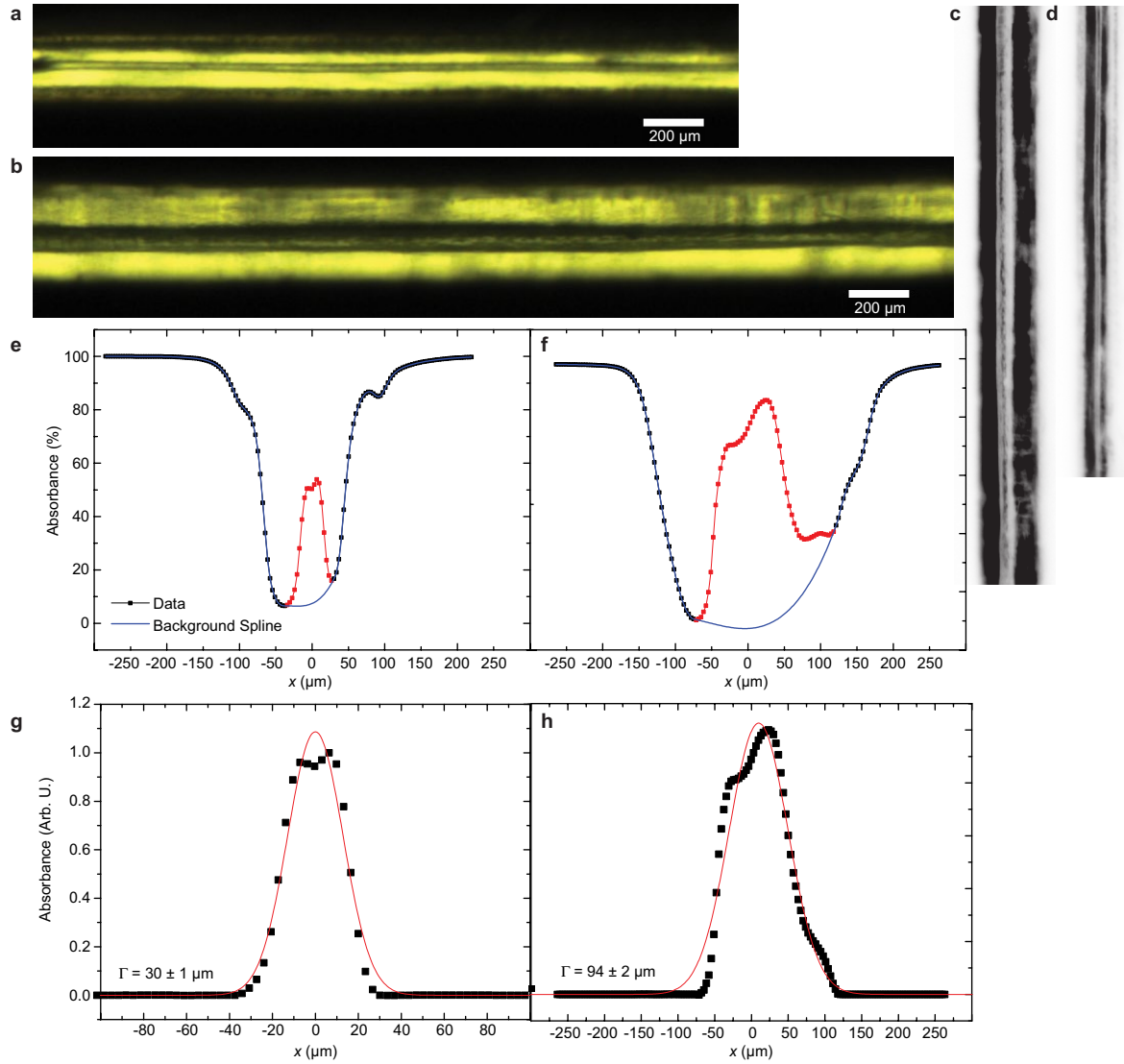
configuration for an antitube (section nearer the viewer) and a pinching region (further section) using magnets ( $6 \times 6 \times 50$  mm) with a gap of 6 mm. **h**, Isosurface plot of the calculated magnetic field for **g**. The weak field region where water can reside is highlighted in white. A water antitube created by the quadrupolar arrangement (left half) is disrupted at the interface between two regions. The field created by the pinching geometry has a field strength of 0.5 T at the centre of magnets (represented by the dark blue colour). **i**, Top view of the Qpump based on this principle. The orientation of magnetization for the arc segments is radially outward (red) or inward (blue). **j**,  $x$ - $y$  contour plot of the calculated magnetic field along the  $z$ -axis centre of the Qpump.



**Extended Data Fig. 7 | Comparison of platelet and whole blood quality in peristaltic and magnetostatic pumping.** **a–c.** Representative SEM images of platelets: after peristaltic pumping (**a**); platelets from control blood (**b**); and platelets from blood after Qpump pumping (**c**). No major morphological

change due to the activation of platelets was observed for either type of pump (peristaltic or Qpump). **d.** A picture of the whole blood in a tube after peristaltic pumping (left), control (middle) and Qpump pumping (right).





**Extended Data Fig. 8 | Process steps for optical imaging of antitubes in Magoil.** **a, b**, Optical micrographs of a water antitube in Magoil for a gap width  $w = 220 \mu\text{m}$  (**a**) and  $w = 307 \mu\text{m}$  (**b**). **c, d**, Greyscale, rotated and inverted images for **b** and **a** respectively. **e, f**, Column average profiles of gap for **a** and **b**, respectively. **g, h**, Gaussian function fits to the background subtracted

profiles. Note: the  $94\text{-}\mu\text{m}$  antitube is thermodynamically stable, as the image was taken during the extrusion of the water antitube, whereas the  $30\text{-}\mu\text{m}$  tube is thermodynamically unstable. After injection, water was then extracted, resulting in a thinning of the tube, which at this diameter collapses into droplets in a matter of minutes.

Extended Data Table 1 | Physical properties of magnetic liquids used

Sample	$\rho$ kg m <sup>-3</sup>	$d_1$ nm	$d_2$ nm	$\phi_1$ %	$\phi_2$ %	$\chi$	$\sigma_{\text{air}}$ mN m <sup>-1</sup>	$\sigma_{\text{water}}$ mN m <sup>-1</sup>	$\sigma_{\text{tween}}$ mN m <sup>-1</sup>
APG 311	950 ± 50	8.2 ± 0.03	—	2.2 ± 4.10 <sup>-3</sup>	—	0.151 ± 0.001	24.7 ± 0.8	22.8 ± 0.8	5.9 ± 0.6
EMG 900	1720 ± 90	11.3 ± 0.1	5 ± 0.1	9.2 ± 0.1	3.3 ± 0.1	1.67 ± 0.04	22 ± 0.5	29.2 ± 1.4	7 ± 0.3
Magoil	970 ± 50	—	—	—	—	4.7×10 <sup>-4</sup> ± 0.5×10 <sup>-4</sup>	—	5.8 ± 0.4	—

Measured physical properties of the two commercial ferrofluids (APG311 and EMG900; Ferrotec) and paramagnetic oil (Magoil): density  $\rho$ , nanoparticle diameters  $d_1$  and  $d_2$  (for bimodal size distributions), corresponding volume fractions  $\phi_1$  and  $\phi_2$  of magnetite in the suspension, initial susceptibility  $\chi$ , surface tension  $\sigma$  with air, surface tension with water, and surface tension with 1% Tween-20 in water.

Extended Data Table 2 | Cell counts of whole blood

	RBC	HGB	HCT	MCV	MCH	MCHC	RDW-SD	RDW-CV	PLT&F	PDW	MPV	P-LCR	PCT	IPF	WBC
unit	10 <sup>6</sup> / μL	g/dL	%	fL	Pg	g/dL	fL	%	10 <sup>3</sup> / μL	fL	fL	%	%	%	10 <sup>3</sup> / μL
ctrl	4.07	12.3	35.5	87.2	30.2	34.6	42.8	13.2	203	11.6	10.5	28.2	0.21	3.3	3.58
Qpump 1	4.11	12.3	35.7	86.9	29.9	34.5	42.1	13.2	169	10.8	10.2	26.8	0.17	3.3	3.44
Qpump 2	3.99	12	34.7	87	30.1	34.6	41.4	13.1	159	11.4	10.5	26.9	0.17	3.3	3.46
Peristaltic pump 1	3.56	10.7	30.4	85.4	30.1	35.2	40.6	13.1	192	11.9	10.7	29	0.22	4.2	3.84
Peristaltic pump 2	4.06	12.3	34.7	85.5											

Comparing a control experiment, after pumping by Qpump, and after peristaltic pumping, measured by haematology analyser (Sysmex XN-1000). Each set of experiments was performed using blood from the same donor. RBC, red blood cells; HGB, haemoglobin; HCT, haematocrit (volume percentage of RBC); MCV, mean corpuscular volume (average RBC volume); MCH, mean corpuscular haemoglobin (average quantity of haemoglobin per RBC); MCHC, mean corpuscular haemoglobin concentration; RDW, RBC distribution width (-SD for standard deviation, -CV for coefficient of variation); PLT&F, platelet counts by fluorescence analysis; PDW, platelet distribution width; MPV, mean platelet volume; P-LCR, platelet large cell ratio; PCT, plateletcrit (volume percentage of platelets in the blood); IPF, immature platelet fraction; WBC, white blood cells.

## Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

### Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- ☒ ☐ The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement
- ☒ ☐ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- ☒ ☐ The statistical test(s) used AND whether they are one- or two-sided  
*Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
- ☒ ☐ A description of all covariates tested
- ☒ ☐ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- ☒ ☐ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- ☒ ☐ For null hypothesis testing, the test statistic (e.g.  $F$ ,  $t$ ,  $r$ ) with confidence intervals, effect sizes, degrees of freedom and  $P$  value noted  
*Give  $P$  values as exact values whenever suitable.*
- ☒ ☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- ☒ ☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- ☒ ☐ Estimates of effect sizes (e.g. Cohen's  $d$ , Pearson's  $r$ ), indicating how they were calculated

*Our web collection on [statistics for biologists](#) contains articles on many of the points above.*

### Software and code

Policy information about [availability of computer code](#)

Data collection

For blood experiments (life sciences): SQUID magnetometer from Quantum Design MPMS3 (software: MPMS 3 MultiVu v 2.3.3.2); Light transmission aggregometry APACT 4004 (software: Apact LPC V 1.19.10); UV-Vis Spectrometry Agilent Cary 8485 (software: Chemstation B.05.04[8]); AAS Agilent 200 Series AAS (software: SpectrAA v 5.4 Pro); Haematology analyser Sysmex Europe XN-1000 (software: XN-Series version 1993200-52); Phenom Pro Desktop SEM (software: SW version 4.3.3)

Data analysis

Origin Pro 9.0, Excel for Mac v.15.33

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

### Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Source data for figure(s) 2, 3, 4 (and extended data figures containing data graphs) are provided with the paper. Any other data that support the findings of this study are available from the corresponding author upon reasonable request.



## Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://nature.com/documents/nr-reporting-summary-flat.pdf)

## Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	For blood pumping studies to determine hemolysis 3 (anonymous) blood donors each donated 50 mL, which was used to make 6 mL samples for each series of experiments (see section M17 for a more detailed description). For studies to determine whether magnetite nanoparticles were transferred from the ferrofluid to the blood during pumping 2 anonymous donors were used.
Data exclusions	Three Qpumps were made (using 3D printing / glued magnet arcs, see section M16). Pump number 3 made loud mechanical noises and its rotation was irregular. All data obtained using pump number 3 was discarded for that reason.
Replication	Blood pumping experiments were done on the same Qpumps, using different blood donors, and on different days/weeks. The results were always consistent.
Randomization	No randomization was used.
Blinding	No blinding was used.

## Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

### Materials & experimental systems

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input type="checkbox"/>	<input checked="" type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data

### Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

## Human research participants

Policy information about [studies involving human research participants](#)

Population characteristics	Blood donors were anonymous, and consist of male / female healthy adults that regularly donate blood to the French blood center (EFS). See ethics oversight below.
Recruitment	A dozen of donors come to the French blood center (EFS) in Strasbourg every day. No specific recruitment or selection was done for this study.
Ethics oversight	Human studies were performed according to the Helsinki Declaration. Control human samples were obtained from volunteer blood donors who gave written informed consent recruited by the blood transfusion center where the research was performed (Etablissement Français du Sang-Grand Est). Patients were recruited at the University Hospital of Strasbourg and signed a statement giving informed consent (no. CODECOH: DC-2015-2368, Comité de Protection des personnes Est-IV no. 15/50 du 15/02/2016) that residual blood collected for classical care can be used for research purposes after irreversible anonymization.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

# Millennial-scale hydroclimate control of tropical soil carbon storage

<https://doi.org/10.1038/s41586-020-2233-9>

Received: 5 September 2018

Accepted: 17 February 2020

Published online: 6 May 2020

 Check for updates

Christopher J. Hein<sup>1✉</sup>, Muhammed Usman<sup>2,3</sup>, Timothy I. Eglinton<sup>2</sup>, Negar Haghipour<sup>2,4</sup> & Valier V. Galy<sup>5</sup>

The storage of organic carbon in the terrestrial biosphere directly affects atmospheric concentrations of carbon dioxide over a wide range of timescales. Within the terrestrial biosphere, the magnitude of carbon storage can vary in response to environmental perturbations such as changing temperature or hydroclimate<sup>1</sup>, potentially generating feedback on the atmospheric inventory of carbon dioxide. Although temperature controls the storage of soil organic carbon at mid and high latitudes<sup>2,3</sup>, hydroclimate may be the dominant driver of soil carbon persistence in the tropics<sup>4,5</sup>; however, the sensitivity of tropical soil carbon turnover to large-scale hydroclimate variability remains poorly understood. Here we show that changes in Indian Summer Monsoon rainfall have controlled the residence time of soil carbon in the Ganges–Brahmaputra basin over the past 18,000 years. Comparison of radiocarbon ages of bulk organic carbon and terrestrial higher-plant biomarkers with co-located palaeohydrological records<sup>6</sup> reveals a negative relationship between monsoon rainfall and soil organic carbon stocks on a millennial timescale. Across the deglaciation period, a depletion of basin-wide soil carbon stocks was triggered by increasing rainfall and associated enhanced soil respiration rates. Our results suggest that future hydroclimate changes in tropical regions are likely to accelerate soil carbon destabilization, further increasing atmospheric carbon dioxide concentrations.

The size of the atmospheric CO<sub>2</sub> reservoir (around 750 Pg C at present) is modulated in part by exchanges with active reservoirs of biospheric organic carbon (BOC), such as marine primary producers, dissolved organic carbon, terrestrial vegetation and soils. Of these, the terrestrial carbon cycle remains the least well constrained<sup>7</sup>, reflecting the complexity of terrestrial ecosystem sensitivity to changes in atmospheric temperature<sup>8,9</sup>, CO<sub>2</sub> concentrations<sup>10</sup>, nutrient availability<sup>11</sup> and hydroclimate<sup>4</sup>. In particular, the sensitivity of soil carbon stocks (with a global reservoir of 1,600 Pg C) to climate remains unresolved, hindering projections of carbon cycling under future climate-change scenarios. Soil carbon storage is a function of inputs from leaf and root detritus associated with net primary productivity; and outputs through lateral export via rivers, CO<sub>2</sub> and CH<sub>4</sub> efflux due to root respiration and microbial decomposition, and fire emissions<sup>1</sup>. Soil carbon respiration rates—and thus turnover times—respond directly to temperature changes<sup>2,3</sup>. However, recent findings have revealed that this effect is muted in the tropics, where hydroclimate variability may have a more important role in climate/carbon-cycle feedback<sup>4,5</sup>. To test this hypothesis, we reconstruct how hydroclimate change since the Late Glacial period (24–18 ka) has affected the dynamics of soil organic carbon cycling in a continent-sized tropical basin. Specifically, we determine the mean age of terrestrial BOC—a proxy for the dynamics of the soil carbon reservoir—eroded from the Ganges–Brahmaputra river basin and stored

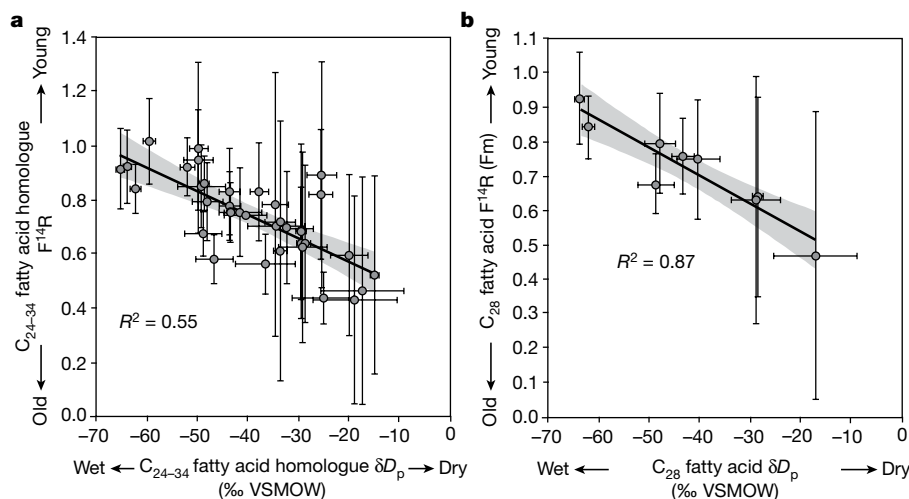
in Bengal Fan sediments. We then compare temporal variations in this mean age with variations in the intensity of the Indian Summer Monsoon (ISM) to reveal the sensitivity of tropical soil carbon stocks to changing hydroclimate.

## Post-glacial climate proxy records

Rivers integrate sediments, organic matter and dissolved constituents from across their drainage basins, allowing downstream archives to capture basin-scale changes in, for example, environmental conditions, sediment fluxes and organic carbon cycling<sup>12</sup>. The radiocarbon (<sup>14</sup>C) content of BOC laterally exported with river sediments provides an estimate of the integrated mean age of BOC in the basin, and, as such, can be used as a proxy for the dynamics of terrestrial organic carbon cycling. Riverine particulate organic carbon is, however, known to be highly heterogeneous<sup>13</sup>: it may include <sup>14</sup>C-free petrogenic carbon derived from rock erosion<sup>14,15</sup> and BOC that is a mixture of fresh organic matter derived from, for example, aquatic productivity and leaf litter, as well as organic matter that is ‘pre-aged’—that is, stored in soils and floodplain deposits before export (for example, see ref. <sup>16</sup>). As such, in addition to bulk organic carbon, we use compound-specific <sup>14</sup>C dating of molecular biomarkers specifically produced by terrestrial vascular plants: long-chain leaf-wax *n*-alkanoic acids (C<sub>24–34</sub> fatty acids).

<sup>1</sup>Department of Physical Sciences, Virginia Institute of Marine Science, William & Mary, Gloucester Point, VA, USA. <sup>2</sup>Geological Institute, Department of Earth Sciences, ETH, Zurich, Switzerland.

<sup>3</sup>Department of Physical & Environmental Sciences, University of Toronto Scarborough, Toronto, Ontario, Canada. <sup>4</sup>Laboratory for Ion Beam Physics, Department of Physics, ETH, Zurich, Switzerland. <sup>5</sup>Department of Marine Chemistry and Geochemistry, Woods Hole Oceanographic Institution, Woods Hole, MA, USA. ✉e-mail: hein@vims.edu



**Fig. 1 | Correlation between palaeoclimate proxies with organic-matter age structure. a, b,** Post-glacial precipitation  $\delta D_p$  values, derived from ice-volume- and vegetation-fractionation-corrected fatty acid  $\delta D_p$  values (see ref. <sup>6</sup> and Methods for details), are plotted against the degree of pre-ageing of organic matter (given as  $F^{14}R$ , in dimensionless units). **a,**  $F^{14}R$  values ( $n = 35$ ) of individual

homologues of  $C_{24-34}$  fatty acids plotted against  $\delta D_p$  values of those same  $C_{24-34}$  fatty acids. **b,**  $F^{14}R$  values ( $n = 9$ ) of  $C_{28}$  fatty acids versus  $\delta D_p$  values of those same  $C_{28}$  fatty acids. Error bars indicate propagated measurement ( $F^{14}R$ ) or standard deviation ( $\delta D_p$ ) errors. VSMOW, Vienna standard mean ocean water;  $R^2$ , coefficient of determination.

To reconstruct variability in soil-carbon storage dynamics within the Ganges–Brahmaputra basin following the Late Glacial, we couple  $^{14}C$  dating of bulk organic carbon and  $C_{24-34}$  fatty acids in fine-grained turbidite beds from a suite of three gravity piston cores collected in the Bengal Fan channel-levee system (see Methods and Extended Data Fig. 1). These cores record basin-scale hydroclimate and vegetation change over a roughly 18,000-year period (Supplementary Table 1 and Extended Data Fig. 2), notably an order-of-magnitude shift in the strength of the ISM between the Late Glacial/Heinrich Stadial 1 (HS1; 18–15 ka) and the Holocene Climatic Optimum (HCO; 10.0–6.5 ka)<sup>6</sup>. Combining these palaeoenvironmental signatures with time-varying lipid age structures from this same suite of biomarkers allows us to quantify the sensitivity of soil-carbon storage within this tropical drainage basin to changing hydroclimatic conditions over millennial timescales.

### Climatic control of organic carbon age structure

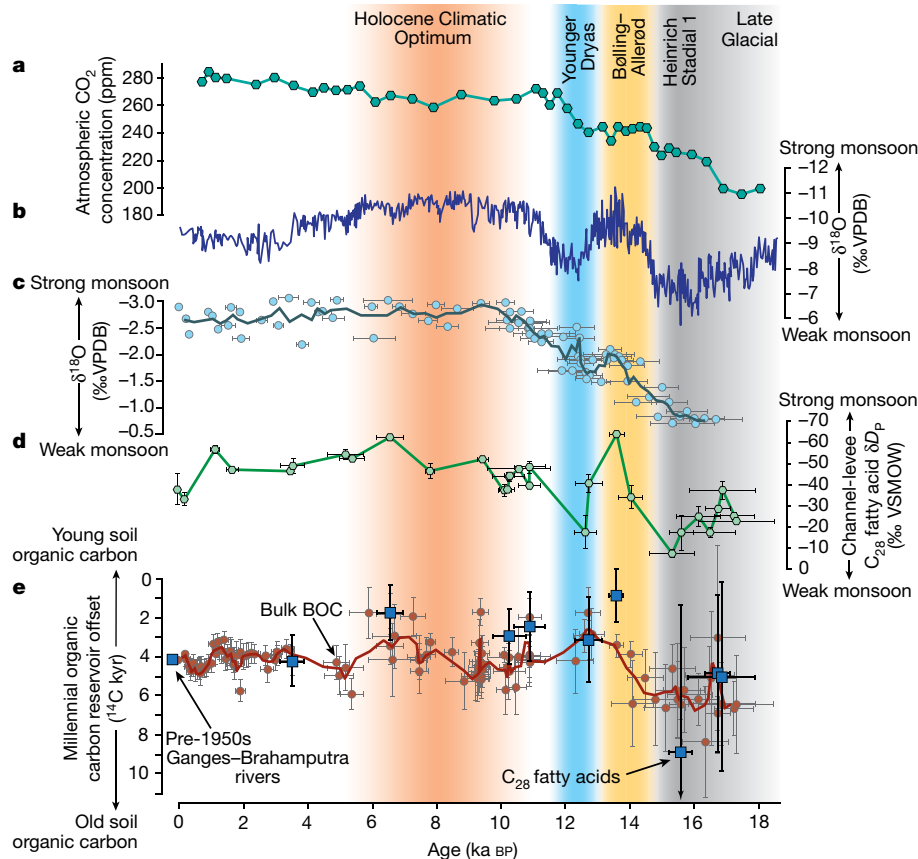
Values for the fraction modern (Fm) of bulk organic carbon (see Methods;  $n = 89$ ) range from 0.085 to 0.648 (Supplementary Table 2). Applying a constant 3% relative contribution of petrogenic organic carbon (following modern values<sup>17</sup>; see Methods), we calculate Fm values for bulk BOC of 0.089–0.683. This large range primarily reflects deposition spanning nearly the past 18,000 years: corrected for deposition age, the radiocarbon age of bulk BOC since the Late Glacial at the time of delivery to the Bengal Fan (its ‘reservoir offset’ value; see Methods and ref. <sup>18</sup>) ranged between around 1,300 and 5,850  $^{14}C$  years (Supplementary Table 2). The mean bulk BOC reservoir offset—3,250  $^{14}C$  years—is nearly identical to those determined for modern Ganges–Brahmaputra river BOC<sup>17</sup>.  $C_{28}$  and weighted-average  $C_{24-32}$  fatty acid homologues have reservoir offset values of around 640–6,150  $^{14}C$  years (average 2,790  $^{14}C$  years) and approximately 375–6,000  $^{14}C$  years (average 2,672  $^{14}C$  years), respectively (Supplementary Table 3 and Extended Data Fig. 3a). We find relatively close agreement between bulk BOC and corresponding  $C_{28}$  and  $C_{24-32}$  fatty acid reservoir offset values; however, bulk BOC for a given sample is systematically more pre-aged (that is, older at the time of export owing to storage on land) than are corresponding long-chain fatty acids (Extended Data Fig. 3b), reflecting the contribution of more diverse, and older, organic matter to the bulk biospheric pool<sup>17</sup>.

Characterized in terms of ‘reservoir relative enrichment’ ( $F^{14}R$ ; see Methods and ref. <sup>18</sup>), the age structures of our bulk BOC and fatty

acids are closely correlated with the hydrogen stable-isotope composition of precipitation ( $\delta D_p$ ) determined from the same compounds<sup>19</sup> (Fig. 1 and Extended Data Fig. 3); weaker correlations with bulk BOC (Extended Data Fig. 4a) probably reflect more heterogeneous source inputs contributing to bulk organic carbon. We can further refine our organic carbon age estimates by taking advantage of the results of ref. <sup>19</sup>, in which numerical modelling was applied to  $^{14}C$  signatures of biomarkers from a sediment core collected from the head of the Bengal Fan channel-levee system, which captures the past 70 years of sediment deposition and was used to estimate the timescales of BOC storage within the modern watershed. They divide the pool of  $C_{24-34}$  fatty acids into ‘fast’ (for example, leaf litter) and ‘slow’ (for example, mineral soil organic carbon) cycling components. These contribute fatty acids cycled on decadal (average age 15 calendar years) and centennial to millennial (average age 1,000–1,200 calendar years) timescales, respectively; the latter constitute 79–83% of the long-chain fatty acid pool. Assuming that fresh, decadal (roughly 15-year cycling time) BOC has always contributed to the Ganges–Brahmaputra rivers from vegetation debris and soil litter at rates similar to today’s (see Methods), we calculate millennial bulk BOC and biomarker ages for the past 18 kyr. We find that, following deglaciation, both closely track—with an apparent lag of roughly 1,000 years—hydroclimate shifts determined from both these same Bengal Fan channel-levee cores and independent records (Fig. 2).

Together, these results indicate that variations in monsoon strength led directly to changes in terrestrial BOC mean age since the Late Glacial. Using  $C_{28}$  fatty acids, for which correlation is the most robust, we find that a 10‰ shift in  $\delta D_p$  values corresponds to a roughly 0.1 change in millennial  $F^{14}R$ . This represents, for example, a change in bulk BOC mean age of 2,600  $^{14}C$  years between HS1 and the HCO, corresponding to a decrease in terrestrial BOC storage times of more than 50%. Given the concurrent transition of terrestrial vegetation from dominantly C4 (savanna-like) to dominantly C3 (forest-like)<sup>6</sup>, which would probably increase overall soil carbon storage capacity, this change in BOC storage time suggests a direct, positive influence of the warmer and wetter conditions within the Ganges–Brahmaputra basin on soil carbon turnover rates.

Given modern-system age distributions<sup>19</sup>, it is not possible to produce  $C_{28} F^{14}R$  values of less than about 0.80, even by increasing the contribution of the millennial BOC component to 100%. Thus, the only means by which to achieve the  $F^{14}R$  values observed throughout the deglacial



**Fig. 2 | Temporal variations in climate forcings and soil organic-carbon storage.** **a–e**, Comparison of: post-glacial wet-extraction atmospheric CO<sub>2</sub> concentrations from Dome Fuji Ice Core<sup>30</sup> using the gas-age profile (**a**); ISM strength as given by a composite Chinese cave speleothem δ<sup>18</sup>O record<sup>31</sup> (**b**), by δ<sup>18</sup>O of planktonic foraminifers (*Globigerinoides ruber*) ( $n = 81$ ) (**c**) and calculated as precipitation δD<sub>p</sub> from terrestrial biomarker δD from the Bengal Fan channel-levee ( $n = 30$ ) (**d**); and records of the age structure of bulk biospheric organic carbon (red circles;  $n = 116$ ) and biomarkers (C<sub>28</sub> fatty acids; blue squares;  $n = 9$ ) exported from the Ganges–Brahmaputra basin (**e**). Data in

**c, d** are replotted from ref. <sup>6</sup>) with our updated core age-depth model. Data in **e** are given as the difference between the age of the millennial component of biomarkers/bulk BOC and the deposition age (in <sup>14</sup>C yrs). Solid lines in **c, e** are three-point moving averages. Pre-1950s Ganges–Brahmaputra river data in **e** are from ref. <sup>19</sup>. Vertical error bars indicate propagated measurement uncertainties (reservoir offset values) or multimeasurement standard deviation (δD) errors for instrumental measurements; horizontal error bars in **c–e** are core age-model deposition-age error estimates (see Methods).

and Holocene period is to increase the age of the millennial component of BOC. This is supported by the high abundance of fatty acids during warm, wet periods (Extended Data Fig. 5) and by moderate correlations between fatty-acid abundance and each bulk and biomarker F<sup>14</sup>R (Extended Data Fig. 6). This reflects the export of biomarker-lean (that is, degraded), old BOC during cooler, drier periods, and implies that our Late Glacial and HS1 millennial BOC age estimates are conservative: given any contribution of decadal BOC during these periods (that is, a decadal-to-millennial BOC ratio of more than 0), then an even older millennial BOC contribution is required to produce such low BOC F<sup>14</sup>R values. Our data thus indicate enhanced (with respect to modern) pre-ageing of BOC within the Ganges–Brahmaputra basin during the cooler, drier periods associated with a weak ISM. The subsequent post-Late Glacial and pre-HCO decreases in the age of the millennial component of the BOC delivered to the Bengal Fan correspond directly to warming and moistening of the basin (Fig. 2).

## Implications for the global carbon cycle

The old apparent age of BOC laterally exported to the Bengal Fan suggests that the modern Ganges–Brahmaputra basin stores and exports large quantities of BOC that accumulated over centuries to millennia<sup>19</sup>. Basin-wide, climate-driven changes in this post-deglaciation age structure track not soil carbon provenance (see Methods), but rather

the residence time of that soil carbon within the terrestrial environment, at least over millennial timescales. Thus, the decrease in millennial BOC mean age from the Late Glacial to post-HS1 periods (Fig. 2e) corresponds to an acceleration in soil carbon turnover (a decreasing mean age) by a factor of nearly two. This would have resulted from enhanced soil organic carbon respiration, providing a positive feedback on atmospheric CO<sub>2</sub> concentrations, with a partial reversal in recent millennia following the HCO.

Considering only the post-Late Glacial period reveals another feature of the observed soil carbon–climate feedback. Given that BOC delivered to the coastal ocean represents only around 3% of total annual soil-carbon turnover within the modern Ganges–Brahmaputra basin (the remaining 97% is attributed to CO<sub>2</sub> fluxes to the atmosphere through heterotrophic respiration and fires<sup>5</sup>), accounting for the observed 4,100 <sup>14</sup>C-year change in millennial BOC age through lateral carbon fluxes (erosion and fluvial export) between HS1 and the Bolling–Allerød, without a commensurate increase in soil respiration fluxes, would require a nearly twofold increase in lateral BOC fluxes. This same change requires only a 7.5% increase in soil respiration fluxes (see Methods). Although our data do not allow us to distinguish between these mechanisms, the dependence of soil organic carbon turnover on hydroclimate is best explained by an imbalance between increasing organic carbon heterotrophic respiration and net primary productivity. The observed 1,000-year lag in millennial reservoir offset



values as compared with hydroclimate proxy records (Fig. 2) probably represents the system response time: an instantaneous shift in soil turnover will take centuries to shift the overall age structure of soil carbon and associated laterally exported organic carbon. Moreover, correlations are much weaker between the millennial age structure of organic carbon and atmospheric CO<sub>2</sub> concentrations (Fig. 2a) than with direct measures of hydroclimate (Fig. 2b–d). This suggests that soil fertilization (increase in productivity related to increasing CO<sub>2</sub>) may not exert a first-order control on soil organic carbon turnover, but rather is filtered through the (hydro)climate system. This implies that, in the tropics, the climate/carbon-cycle feedback is not simple and one-directional, but rather is responsive to the impact of CO<sub>2</sub>-associated warming on terrestrial precipitation patterns, confirming at the millennial scale, and in a globally important system, observations from subdecadal-scale arid-climate macrocosm experiments<sup>20</sup>.

The Ganges–Brahmaputra rivers are globally important in terms of sediment load ( $1\text{--}2 \times 10^9 \text{ t yr}^{-1}$ ; ref. <sup>21</sup>), carbon burial (roughly  $4.7 \times 10^{12} \text{ g yr}^{-1}$ , around 20% of the global terrestrial BOC burial flux; refs. <sup>22,23</sup>), and marine sequestration of terrigenous organic carbon at timescales of thousands to millions of years<sup>6,22,24</sup>. Our results demonstrate that warmer, wetter conditions associated with post-glacial strengthening of the ISM led to release of pre-aged BOC from within this basin. Although it is not clear that increased atmospheric CO<sub>2</sub> concentrations (and associated climatic warming) directly lead to increased ISM rainfall intensity<sup>25</sup>, we find an apparent positive feedback between hydroclimate and atmospheric CO<sub>2</sub> concentrations, acting through changes in the soil carbon stock over millennial timescales, most likely triggered by enhanced heterotrophic respiration. In this manner, we find a low-latitude climate feedback similar to that reported for high-latitude permafrost soils<sup>26</sup>. This supports findings from carbon-cycle models of glacial gross primary productivity and terrestrial carbon stocks, which suggest overall turnover times and associated large ‘inert’ terrestrial carbon stocks (in soils, permafrost and wetlands) nearly doubled during the last glacial period<sup>27</sup>. Our study reveals that tropical soils may have contributed substantially to this glacial ‘inert’ terrestrial carbon pool—previously attributed nearly exclusively to permafrost<sup>27</sup>. Our results further show that future hydroclimate change—in which decreases in the strength of the ISM are likely to be more than compensated by increased atmospheric moisture content, leading to more overall precipitation<sup>28,29</sup>—is likely to further accelerate soil organic carbon destabilization (turnover rates and/or net loss of carbon stocks) in the tropics, increasing atmospheric BOC fluxes through enhanced soil decomposition or transport-induced oxidation, and thus further increasing atmospheric CO<sub>2</sub> concentrations.

## Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-020-2233-9>.

- Davidson, E. A. & Janssens, I. A. Temperature sensitivity of soil carbon decomposition and feedbacks to climate change. *Nature* **440**, 165–173 (2006).

- Melillo, J. M. et al. Soil warming and carbon-cycle feedbacks to the climate system. *Science* **298**, 2173–2176 (2002).
- Lloyd, J. & Taylor, J. A. On the temperature dependence of soil respiration. *Funct. Ecol.* **8**, 315–323 (1994).
- Carvalho, N. et al. Global covariation of carbon turnover times with climate in terrestrial ecosystems. *Nature* **514**, 213–217 (2014).
- Bloom, A. A., Exbrayat, J. F., van der Velde, I. R., Feng, L. & Williams, M. The decadal state of the terrestrial carbon cycle: global retrievals of terrestrial carbon allocation, pools, and residence times. *Proc. Natl Acad. Sci. USA* **113**, 1285–1290 (2016).
- Hein, C. J., Galy, V. V., Galy, A., France-Lanord, C. & Kudrass, H. Post-glacial climate forcing of surface processes in the Ganges–Brahmaputra river basin and implications for carbon sequestration. *Earth Planet. Sci. Lett.* **478**, 89–101 (2017).
- Friend, A. D. et al. Carbon residence time dominates uncertainty in terrestrial vegetation responses to future climate and atmospheric CO<sub>2</sub>. *Proc. Natl Acad. Sci. USA* **111**, 3280–3285 (2014).
- Xia, J. et al. Terrestrial carbon cycle affected by non-uniform climate warming. *Nat. Geosci.* **7**, 173–180 (2014).
- Hicks Pries, C. E. H., Castanha, C., Porras, R. C. & Torn, M. S. The whole-soil carbon flux in response to warming. *Science* **355**, 1420–1423 (2017).
- Schimmel, D., Stephens, B. B. & Fisher, J. B. Effect of increasing CO<sub>2</sub> on the terrestrial carbon cycle. *Proc. Natl Acad. Sci. USA* **112**, 436–441 (2015).
- Wieder, W. R., Cleveland, C. C., Smith, W. K. & Todd-Brown, K. Future productivity and carbon storage limited by terrestrial nutrient availability. *Nat. Geosci.* **8**, 441–444 (2015).
- Simpson, G. & Castelltort, S. Model shows that rivers transmit high-frequency climate cycles to the sedimentary record. *Geology* **40**, 1131–1134 (2012).
- Aufdenkampe, A. K. et al. Organic matter in the Peruvian headwaters of the Amazon: compositional evolution from the Andes to the lowland Amazon mainstem. *Org. Geochem.* **38**, 337–364 (2007).
- Drenzek, N. et al. A new look at old carbon in active margin sediments. *Geology* **37**, 239–242 (2009).
- Hilton, R. G., Galy, A., Hovius, N. & Horng, M. J. Efficient transport of fossil organic carbon to the ocean by steep mountain rivers: an orogenic carbon sequestration mechanism. *Geology* **39**, 71–74 (2011).
- Schefeuf, E. et al. Hydrologic control of carbon cycling and aged carbon discharge in the Congo River basin. *Nat. Geosci.* **9**, 687–690 (2016).
- Galy, V. & Eglinton, T. I. Protracted storage of biospheric carbon in the Ganges–Brahmaputra basin. *Nat. Geosci.* **4**, 843–847 (2011).
- Soulet, G., Skinner, L. C., Beaupré, S. R. & Galy, V. A note on reporting of reservoir <sup>14</sup>C disequilibrium and age offsets. *Radiocarbon* **58**, 205–211 (2016).
- French, K. L. et al. Millennial soil retention of terrestrial organic carbon matter deposited in the Bengal Fan. *Sci. Rep.* **8**, 11997 (2018).
- Yu, H. et al. Soil carbon release responses to long-term versus short-term climatic warming in an arid ecosystem. *Biogeosci.* **17**, 781–792 (2020).
- Milliman, J. D. & Syvitski, J. P. Geomorphic/tectonic control of sediment discharge to the ocean: the importance of small mountainous rivers. *J. Geol.* **100**, 525–544 (1992).
- Galy, V. et al. Efficient organic carbon burial in the Bengal fan sustained by the Himalayan erosional system. *Nature* **450**, 407–410 (2007).
- Galy, V., Peucker-Ehrenbrink, B. & Eglinton, T. Global carbon export from the terrestrial biosphere controlled by erosion. *Nature* **521**, 204–207 (2015).
- France-Lanord, C. & Derry, L. A. Organic carbon burial forcing of the carbon cycle from Himalayan erosion. *Nature* **390**, 65–67 (1997).
- Roxy, M. K. et al. Drying of Indian subcontinent by rapid Indian Ocean warming and a weakening land-sea thermal gradient. *Nat. Commun.* **6**, 7423 (2015).
- Gustafsson, Ö., Van Dongen, B. E., Vonk, J. E., Dudarev, O. V. & Semiletov, I. P. Widespread release of old carbon across the Siberian Arctic echoed by its large rivers. *Biogeosciences* **8**, 1737–1743 (2011).
- Ciais, P. et al. Large inert carbon pool in the terrestrial biosphere during the Last Glacial Maximum. *Nat. Geosci.* **5**, 74–79 (2012).
- Christensen, J. H. et al. in *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (eds Stocker, T. F. et al.) 1217–1308 (Cambridge University Press, 2013).
- Sharmila, S., Joseph, S., Sahai, A. K., Abhilash, S. & Chattopadhyay, R. Future projection of Indian summer monsoon variability under climate change scenario: an assessment from CMIP5 climate models. *Global Planet. Change* **124**, 62–78 (2015).
- Kawamura, K. et al. Northern Hemisphere forcing of climatic cycles in Antarctica over the past 360,000 years. *Nature* **448**, 912–916 (2007).
- Cheng, H. et al. The Asian monsoon over the past 640,000 years and ice age terminations. *Nature* **534**, 640–646 (2016); corrigendum **541**, 122 (2016).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2020

## Methods

### Sediment cores

Three gravity piston cores were retrieved at water depths of 2,540–2,610 m along the middle fan channel-levee system (inner-levee, outer-levee, and levee-proximal) during the February 1994 (Expedition SO93) cruise of the R/V SONNE<sup>32</sup>, coincident with sediment echosounder data (Extended Data Fig. 1; see ref. <sup>6</sup> and its associated supplementary information for details). Sediment core age models are updated from ref. <sup>6</sup> and based on a recalibration of 22 reported<sup>6,32</sup> radiocarbon (<sup>14</sup>C) ages derived from planktonic foraminifera (*Globigerinoides sacculifer* and *Globigerinoides ruber*) and accelerator mass spectrometry (AMS) radiocarbon dating of planktonic *G. ruber* from an additional four sedimentary units within core SO93-118KL (Supplementary Table 1). Radiocarbon ages were calibrated using OxCal 4.2 (ref. <sup>33</sup>) with the IntCal13 curve<sup>34</sup> and a regionally averaged marine reservoir correction ( $\Delta R$ ) of  $-4 \pm 42$  years (data from refs. <sup>35,36</sup>). Age models were calculated in the age-depth modelling software package Clam 2.1 (ref. <sup>37</sup>) with extrapolation to core bottoms (Extended Data Fig. 2). Together, these capture the last 17.5 kyr of deposition, with linearized sediment accumulation rates of less than 50 cm kyr<sup>-1</sup> to more than 500 cm kyr<sup>-1</sup> (see ref. <sup>6</sup> for details).

### Bulk organic carbon radiocarbon analysis

Following fumigation acidification, radiocarbon analysis was completed on powdered bulk-sediment aliquots from 89 samples (each representing 1–10 cm thick core sections). Samples were fumigated by sealing in a vacuum desiccator with a beaker of 50 ml 12N HCl, exposed to HCl fumes for 60–72 h at 60–65 °C to remove carbonates, and dried in a separate desiccator for an additional 24 h. Radiocarbon analyses were conducted at either the National Ocean Science Accelerator Mass Spectrometry (NOSAMS) facility (Woods Hole, MA) or with the ETH Zurich AMS 'MICADAS' system (see ref. <sup>38</sup> for details). Detailed results are provided in Supplementary Table 2.

### Compound-specific radiocarbon analysis

Fatty acids derived from lipid extracts from a subset of nine samples were purified by aminopropyl-silica-gel column chromatography, derivatized with methanol of known isotopic composition to produce fatty acid methyl esters (FAMES), and then further purified by aminopropyl-silica-gel chromatography and silver nitrate-silica-gel chromatography (see ref. <sup>6</sup> for details). Six individual, saturated, long-chain FAMES (*n*-C<sub>24</sub>, *n*-C<sub>26</sub>, *n*-C<sub>28</sub>, *n*-C<sub>30</sub>, *n*-C<sub>32</sub> and *n*-C<sub>34</sub>) were purified and collected according to the preparative capillary gas chromatography (PCGC) method<sup>17,39</sup>. This was done through a suite of up to 100, 3–5 µl injections of a saturated FAME solution in iso-octane (individual homologues at 50–200 ng µl<sup>-1</sup> concentrations) on either an Agilent 7890A or an HP 5890 Series II gas chromatograph (GC) coupled to a Gerstel preparative fraction collector (PFC). PCGC traps were recovered with 4 ml of DCM and further purified to remove column bleed by 1% deactivated silica-gel column chromatography, using around 3 cm of gel and eluting FAME homologues in 4 ml of DCM. Homologue purity was checked by injecting a small aliquot on a GC-flame ionization detector (FID), and yields were determined to a range between around 40% and 80%. To increase sample size and reduce analytical uncertainty during radiocarbon analyses, it was necessary to combine homologues following isolation by PCGC for two samples: (1) for sample SO93-120KL 1,130–1,140 cm, two sets of homologues were combined: *n*-C<sub>24</sub> and *n*-C<sub>26</sub> (treated as a C<sub>24</sub> sample for purposes of data analysis) and *n*-C<sub>28</sub>, *n*-C<sub>30</sub> and *n*-C<sub>32</sub> (treated as a C<sub>28</sub> sample); and (2) for sample SO93-120KL 632–642 cm, *n*-C<sub>32</sub> and *n*-C<sub>34</sub> homologues were combined (treated as a C<sub>32</sub> sample).

PCGC-purified FAME homologues were dissolved in DCM and loaded into premuffled quartz tubes. Solvent was evaporated under a high-purity nitrogen stream and combusted copper oxide (roughly

150 µg) was added to the quartz tube. Quartz tubes were evacuated to less than 30 µTorr and flame-sealed on a vacuum line following chilling using dry ice/isopropanol slurry and liquid nitrogen cryogenic traps to prevent sublimation of FAMES. Samples were combusted in flame-sealed quartz tubes at 850 °C for 5 h, and, the following day, were cracked under vacuum; the released CO<sub>2</sub> was cryogenically purified and manometrically quantified. The sample CO<sub>2</sub> was trapped using liquid nitrogen, flame-sealed in a pyrex tube, and sent to either ETH Zurich or NOSAMS where they were analysed for <sup>14</sup>C content as a gas target with the Micadas gas ion source as described<sup>38</sup> (ETH Zurich), or as a graphite target following catalytic conversion as described<sup>40,41</sup> (NOSAMS).

Resulting <sup>14</sup>C measurements were corrected for procedural blanks (the blank correction varies among the samples, depending on the mass of the measured fractions; see Supplementary Table 3), and for the addition of derivative carbon during methylation with MeOH with a known <sup>14</sup>C composition, using mass-balance equations (see ref. <sup>19</sup> for details).

A total of 40 <sup>14</sup>C measurements were made on homologues (or combined homologues) from the 9 samples selected for analysis. Not all homologues were measured from all samples owing to insufficient compound availability or loss during sample preparation (for example, shattering of quartz tubes). Radiocarbon data are expressed as the fraction of modern carbon, Fm, precorrected for mass-dependent fractionation using the <sup>13</sup>C/<sup>12</sup>C ratio measured online.

We observed relatively small differences in stable-isotope ( $\delta D$ )<sup>6</sup> and radioisotope (Extended Data Fig. 3a and Supplementary Table 3) compositions among C<sub>24–34</sub> fatty acid homologues. Fatty acids generally, although not universally, become more <sup>14</sup>C-depleted with increasing chain length, a phenomenon commonly associated with the selective preservation of long-chain homologues<sup>42,43</sup>. C<sub>28</sub> fatty acids—which were consistently the most abundant homologue (average concentration 0.64 µg g<sup>-1</sup>) and yielded the highest average C mass for <sup>14</sup>C analysis (48.8 µg)—and weighted-average C<sub>24–32</sub> fatty acids are the focus of discussion.

### <sup>14</sup>C deposition ages

Sediment deposition ages (in <sup>14</sup>C years) are determined from <sup>14</sup>C analysis of planktonic foraminifera (*G. sacculifer* and *G. ruber*; Supplementary Table 1), corrected for a standard 400-year surface ocean <sup>14</sup>C reservoir age offset. Deposition of 15 of our 118 samples is directly dated; the remaining deposition ages (with errors) are calculated through interpolation from <sup>14</sup>C age-depth models constructed in the Clam 2.1 software package<sup>37</sup>.

### Radiocarbon terminology, conventions and calculations

Sample <sup>14</sup>C contents are given in the nondimensional 'fraction modern' (Fm) notation—a measurement of the deviation of a sample <sup>14</sup>C/<sup>12</sup>C ratio from 95% of the radiocarbon concentration at 'modern' (1950 CE) of NBS oxalic acid I (SRM 4990B, OX-I), normalized to a  $\delta^{13}C_{VPDB}$  value of  $-19\text{‰}$  (see refs. <sup>44–46</sup>). Given that Fm is a measure of the <sup>14</sup>C/<sup>12</sup>C ratio, it is, by definition, a measure of the radiocarbon content of a sample, and can thus be related to the radiocarbon (uncalibrated) age of the sample as follows:

$$^{14}\text{C age} = -8033 \times \ln(\text{Fm})$$

Here, we seek to determine the amount of 'pre-aging' that organic matter has undergone before export and delivery to the coastal ocean, as a means of estimating soil organic carbon turnover rates. To accomplish this, we use the 'reservoir offset' metric, defined<sup>18</sup> as a measure of the radiocarbon age offset, in <sup>14</sup>C years, between two contemporaneous carbon reservoirs. This metric relies on the 'relative enrichment' of the reservoir, defined<sup>18</sup> as the ratio of Fm values from two contemporaneous carbon reservoirs (*x* and *y*) at a given time (in our case, the time of deposition in the Bengal Fan). It is thus a measure of the age

(in  $^{14}\text{C}$  years) of organic matter at the time of its deposition. Relative enrichment is given by the notation  $F^{14}\text{R}$  in dimensionless units, as follows:

$$F^{14}\text{R}_{x-y} = \frac{F_{m_x}}{F_{m_y}}$$

By convention,  $F^{14}\text{R}$  is dimensionless and ranges from 0 to 1 by placing the more commonly enriched reservoir (y) in the denominator. From this, the reservoir age offset ( $R_{x-y}$ ) can be determined:

$$R_{x-y} = -8033 \times \ln(F^{14}\text{R}_{x-y})$$

In order to calculate reservoir age offset values through time, it is necessary to ‘uncalibrate’ organic-matter deposition ages derived from the core-age models (derived from calibrated  $^{14}\text{C}$  ages of foraminifera, and plotted in calibrated (or ‘calendar’) years before present (BP), where ‘present’ is set by convention to the year 1950 CE). For this, we take advantage of the Bayesian ‘uncalibration-convolution process’ approach of ref. <sup>47</sup>, which propagates uncertainties linked to the reservoir-derived  $^{14}\text{C}$  age (in this case, bulk organic matter or biomarker  $F_m$  values), a weakly a priori known calendar age (in this case, derived from core-age models), and the atmospheric calibration curve (from IntCal13; ref. <sup>34</sup>). Specifically, we apply to our biomarker, bulk organic carbon, and bulk biospheric organic carbon ages the *radcal* script of the ResAge Package<sup>47</sup>, designed for pairs of reservoir-derived radiocarbon ages and a corresponding weakly known (with some uncertainty) calendar ages. This program uncalibrates age-model-derived deposition ages (that is, it converts these to  $^{14}\text{C}$  years) and calculates reservoir offset and  $F^{14}\text{R}$  values for each input.

## Determination of millennial BOC ages

$F^{14}\text{R}$  values for each bulk and fatty acid homologue sample can be expressed as a function of the age (in  $F_m$ ) of each the millennial (mil) and decadal (dec) pools, multiplied by the fractional contribution of each of those pools.

We hold the contributions to the overall Ganges–Brahmaputra organic-carbon pool of each of the decadal BOC ( $F_{\text{dec}}$ ), millennial BOC ( $F_{\text{mil}}$ ), and petrogenic organic carbon ( $F_{\text{rxC}}$ ) constant through time. Specifically, petrogenic carbon is assumed to contribute 3% of all organic carbon to the bulk carbon pool (BOC is roughly 97% of total organic carbon), whereas fatty acids are free of petrogenic carbon (BOC roughly 100%); this assumption is founded in analyses of each river and Bengal Fan core-top sediments, which reveal that Ganges–Brahmaputra river sediments delivered to the Bay of Bengal at present contain roughly 0.03% petrogenic organic carbon<sup>17,48,49</sup>. In addition, structural characterization (Raman and high-resolution transmission electron microscopy) of petrogenic organic carbon in sediments from the Bengal Fan dating back to the Miocene<sup>48</sup>, coupled with evidence<sup>6</sup> for high burial efficiency of terrestrial organic carbon over the past 20,000 years, suggests that the petrogenic organic-carbon concentration in Bengal Fan sediments has remained constant since the Last Glacial. Finally, we would anticipate that, if petrogenic contributions changed in any systematic manner through time, it would be reflected in bulk total organic carbon (TOC) values; indeed, bulk TOC shows no correlation either with the relative enrichment of bulk organic-carbon radiocarbon ( $R^2 = 0.0036$ ; Extended Data Fig. 7a), or with petrogenic-contribution-corrected BOC relative enrichment ( $R^2 = 0.0025$ ) or millennial BOC relative enrichment ( $R^2 = 0.0024$ ).

Given these constraints, we apply time-invariant millennial and decadal contributions to BOC of 80% and 20%, respectively (values for modern Ganges–Brahmaputra rivers are from ref. <sup>19</sup>). The fatty-acid decadal pool is assumed to always turnover quickly (by definition), and is assigned a constant 15-year cycling time ( $F^{14}\text{R} = 0.9981$ )<sup>19</sup>. The  $^{14}\text{C}$  age of the millennial BOC is calculated at each time step as follows:

$$F^{14}\text{R}_{\text{tot}} = F^{14}\text{R}_{\text{mil}} \times F_{\text{mil}} + F^{14}\text{R}_{\text{dec}} \times F_{\text{dec}} + F^{14}\text{R}_{\text{rxC}} \times F_{\text{rxC}}$$

## Constraining grain-size and palaeovegetation effects

Sediment organic-carbon concentrations are highly sensitive to sediment grain size, which, within the Ganges–Brahmaputra rivers and Bengal Fan, shows close correspondence with the production of aluminosilicate clays, and thus sediment Al/Si content<sup>6,22,48,50</sup>. However, neither the grain size (Al/Si) nor the grain-size-normalized TOC values of the Bengal Fan channel-levee cores show any discernible trend through time<sup>6</sup>. Further, we find no correlation between organic-matter age with respect to deposition time (that is,  $F^{14}\text{R}$  values) and either TOC or Al/Si values (Extended Data Fig. 7).

Changes in vegetation since the Last Glacial were documented in ref. <sup>6</sup> using bulk and biomarker  $\delta^{13}\text{C}$  values from the same cores used here. These changes were accounted for in ref. <sup>6</sup> in the development of the biomarker  $\delta D_p$  record (derived from the lipid  $\delta D$  record); we directly used those same biomarker-derived  $\delta D_p$  values here: a subset is presented in each of Fig. 1 and Extended Data Fig. 4, and the full  $\delta D_p$  record (that is, lipid  $\delta D$  values corrected for ice volume and palaeovegetation (from lipid  $\delta^{13}\text{C}$  values)) is given in Fig. 2.

Corrections of lipid  $\delta D$  values for ice-volume and lipid vegetation are described in full in ref. <sup>6</sup>, and summarized briefly as follows. Methanol-corrected  $\delta D$  values from fatty-acid methyl esters ( $\delta D_{\text{FA}}$ ) were corrected for variations in seawater isotope composition related to global ice-volume variations during the deglaciation using a record of global relative sea-level changes<sup>51</sup> to infer variations in the oxygen-isotope composition of sea water ( $\delta^{18}\text{O}_{\text{SW}}$ ) related to variations in global ice volume across the deglaciation. Relative sea-level variations were converted to variations in  $\delta^{18}\text{O}_{\text{SW}}$  using a scaling factor of 1.0‰ per 127.5 m of relative sea-level change<sup>52</sup>. Variations in  $\delta^{18}\text{O}_{\text{SW}}$  were subsequently converted to variations in  $\delta D_{\text{SW}}$  using the slope of the global meteoric water line. Finally, ice-volume-corrected FAMES  $\delta D$  values ( $\delta D_{\text{FA-IV}}$ ) were calculated by correcting measured  $\delta D_{\text{FA}}$  for changes in  $\delta D_{\text{SW}}$  relative to the modern seawater composition.  $\delta D_{\text{FA-IV}}$  values were then corrected for variable D/H fractionation by each of C3 and C4 vegetation endmembers to estimate precipitation  $\delta D$  values ( $\delta D_p$ ). Estimated D/H fractionation factors ( $\epsilon$ ) of  $-125\text{‰}$  (100% C3 trees vegetation;  $\epsilon_{\text{C3}}$ ) and  $-145\text{‰}$  (100% C4 grasses vegetation;  $\epsilon_{\text{C4}}$ ) were applied on the basis of variable C3 and C4 mixing proportions determined from sample FAMES  $\delta^{13}\text{C}$  values following ref. <sup>53</sup>.

## Constraining soil carbon fluxes

Under steady-state conditions, the soil carbon turnover time ( $T_r$ ) can be given as a function of the total mass of soil carbon ( $M_s$ ) and soil carbon inputs (NPP ( $F_p$ )) and outputs ( $F_{\text{out}}$ ), where  $F_{\text{out}}$  is a combination of fluxes from heterotrophic respiration ( $F_r$ ), fire emissions ( $F_f$ ) and river erosion and export (lateral fluxes,  $F_l$ ) as follows:

$$T_r = \frac{M_s}{F_p} = \frac{M_s}{F_r + F_f + F_l} \quad (1)$$

We first assume that, since the Last Glacial,  $F_{\text{out}}$  responds only to changes in  $F_r$  and  $F_f$ . Following the Last Glacial, and through the HCO, climate in the Ganges–Brahmaputra basin gradually becomes wetter, and vegetation trends towards greater dominance by C3 forests, as opposed to C4 savanna<sup>6</sup>. Together, these would be far more likely to decrease, as opposed to increase, fire intensity, and thus fire carbon emissions. Thus, holding  $F_r$  constant presents a conservative estimate of changes in  $M_s$  and/or  $F_r$  and  $F_l$  between the Last Glacial and post-HS1 periods. We also assume that  $F_r$  and  $F_l$  remain at their present relative proportions of 97% and 3%, respectively, of the non- $F_f$  portion of  $F_{\text{out}}$ . Finally, we assume that soil (millennial) carbon always composes 80% of total carbon burial (see text for details) and that carbon burial efficiency (sediment-volume and grain-size-normalized

burial rates) have remained near 100% since the Last Glacial (discussed in ref. <sup>6</sup>).

Our bulk organic-carbon radiocarbon results indicate that the reservoir-age offset (proxy for  $T_r$ ) of millennial (soil) BOC pools decreased by around 3,800  $^{14}\text{C}$  years (approximately 56%) between HS1 and the HCO (Supplementary Table 2). Although changing either  $F_p$  or  $F_{\text{out}}$  would affect the steady-state turnover time (as per equation (1)), it has no immediate impact on the average age of soil organic carbon, and thus the average BOC reservoir offset value. Thus, observed post-Last Glacial shifts in laterally fluxed BOC are best explained by changes in Ganges–Brahmaputra soil carbon stocks over this time (that is, time-varying  $M_s$ ). For example, holding  $F_p$  and  $F_{\text{out}}$  constant, we would expect that  $M_s$  would vary between approximately 33.2 Pg C at HS1, to roughly 14.4 Pg C at the HCO (a depletion of soil carbon stocks of roughly 56%), and to its modern documented value of 19.52 Pg C (see ref. <sup>4</sup>). Likewise, between HS1 and the post-Bølling–Allerød millennial organic-carbon reservoir offset minimum (around 2,500  $^{14}\text{C}$  years), basin-wide carbon stocks decreased by an amount (about 20.9 Pg) nearly equivalent to the modern basin-wide storage.

Given that BOC delivered to the coastal ocean represents only around 3% of total annual soil-carbon turnover within the Ganges–Brahmaputra basin (about 0.0036 Pg yr<sup>-1</sup>; refs. <sup>23,49</sup>), the HS1 to post-Bølling–Allerød decrease in basin-wide soil carbon stocks, if driven entirely by changes in river carbon export, would require lateral fluxes to nearly double for this roughly 3,000-year period. By contrast, a mere 7.5% increase in heterotrophic respiration rates would be required to account for the observed decrease in soil carbon stocks between HS1 and the Bølling–Allerød.

### Constraining impacts of soil carbon provenance

Changes in the age structure of BOC laterally exported from the Ganges–Brahmaputra rivers cannot be explained by sediment-source variability from within-basin soil profiles (for example, because of erosion of deeper, older carbon during cooler and drier periods, potentially in response to reduced vegetation cover). Since deglaciation, sediments exported by the Ganges–Brahmaputra rivers have become gradually more chemically weathered<sup>6,50</sup>, whereas the BOC exported with these same sediments has become younger; thus, sediment ‘age’ (as given by pre-export weathering) and BOC age structures are decoupled. Changes in BOC age structure also cannot be explained by changes in sediment provenance from within the basin (for example, by shifting foci of monsoon precipitation associated with changes in monsoon strength<sup>6</sup>). Whereas sediment may be sourced from throughout the Ganges–Brahmaputra basin, BOC delivered to the Bengal Fan is predominantly floodplain in origin, as upland organic carbon is subject to efficient degradation and replacement during transit<sup>54</sup>. BOC age structures should thus be insensitive to sediment sourcing upstream of the floodplain.

### Data availability

All new data produced for this study are from samples from Bengal Fan cores SO93-117KL, -118KL and -120KL. These are available in Supplementary Tables and online in the EarthChem Library (<http://www.earthchem.org/>). Specifically, these present a compilation of radiocarbon age-dating results from planktonic foraminifera used in derivation of core age-depth models (Supplementary Table 1 and <https://doi.org/10.1594/IEDA/111486>); results of radiocarbon analyses of bulk organic carbon and calculated reservoir offset and  $F^{14}\text{R}$  values of bulk and millennial BOC (Supplementary Table 2 and <https://doi.org/10.1594/IEDA/111487>); results of radiocarbon analyses of fatty-acid homologues and associated calculated reservoir offset and  $F^{14}\text{R}$  values of bulk homologues and the subset of those cycled on millennial

timescales (Supplementary Table 3 and <https://doi.org/10.1594/IEDA/111488>); and abundances of fatty-acid homologues (Supplementary Table 4 and <https://doi.org/10.1594/IEDA/111489>). Source data for Figs. 1, 2 and Extended Data Figs. 2–7 are provided with the paper.

32. Weber, M. E., Wiedicke, M. H., Kudrass, H. R., Hübscher, C. & Erlenkeuser, H. Active growth of the Bengal Fan during sea-level rise and highstand. *Geology* **25**, 315–318 (1997).
33. Bronk Ramsey, C. B. Bayesian analysis of radiocarbon dates. *Radiocarbon* **51**, 337–360 (2009).
34. Reimer, P. J. et al. IntCal13 and Marine13 radiocarbon age calibration curves 0–50,000 years cal BP. *Radiocarbon* **55**, 1869–1887 (2013).
35. Dutta, K., Bhushan, R. & Somayajulu, B.  $\Delta R$  correction values for the northern Indian Ocean. *Radiocarbon* **43**, 483–488 (2001).
36. Southon, J., Kashgarian, M., Fontugne, M., Metivier, B. & Yim, W. W. Marine reservoir corrections for the Indian Ocean and Southeast Asia. *Radiocarbon* **44**, 167–180 (2002).
37. Blaauw, M. Methods and code for ‘classical’ age-modelling of radiocarbon sequences. *Quat. Geochronol.* **5**, 512–518 (2010).
38. Christl, M. et al. The ETH Zurich AMS facilities: performance parameters and reference materials. *Nucl. Instrum. Methods Phys. Res. B* **294**, 29–38 (2013).
39. Eglinton, T. I. et al. Gas chromatographic isolation of individual compounds from complex matrices for radiocarbon dating. *Anal. Chem.* **68**, 904–912 (1996).
40. Santos, G. M. et al. Blank assessment for ultra-small radiocarbon samples: chemical extraction and separation versus AMS. *Radiocarbon* **52**, 1322–1335 (2010).
41. Shah Walter, S. R. et al. Ultra-small graphitization reactors for ultra-microscale  $^{14}\text{C}$  analysis at the National Ocean Sciences Accelerator Mass Spectrometry (NOSAMS) facility. *Radiocarbon* **57**, 109–122 (2015).
42. Kusch, S., Rethemeyer, J., Schefuß, E. & Mollenhauer, G. Controls on the age of vascular plant biomarkers in Black Sea sediments. *Geochim. Cosmochim. Acta* **74**, 7031–7047 (2010).
43. Ohkouchi, N., Eglinton, T. I. & Hayes, J. M. Radiocarbon dating of individual fatty acids as a tool for refining Antarctic margin sediment chronologies. *Radiocarbon* **45**, 17–24 (2003).
44. Olsson, I. U. (ed). *Radiocarbon Variations and Absolute Chronology: Nobel Symposium, 12th Proc.* (John Wiley, 1970).
45. Stuiver, M. & Polach, H. A. Discussion: reporting of  $^{14}\text{C}$  data. *Radiocarbon* **19**, 355–363 (1977).
46. Reimer, P. J., Brown, T. A. & Reimer, R. W. Discussion: reporting and calibration of post-bomb  $^{14}\text{C}$  data. *Radiocarbon* **46**, 1299–1304 (2004).
47. Soulet, G. Methods and codes for reservoir-atmosphere  $^{14}\text{C}$  age offset calculations. *Quat. Geochronol.* **29**, 97–103 (2015).
48. Galy, V., Beyssac, O., France-Lanord, C. & Eglinton, T. I. Recycling of graphite during Himalayan erosion: a geological stabilization of carbon in the crust. *Science* **322**, 943–945 (2008).
49. Galy, V., Hein, C., France-Lanord, C. & Eglinton, T. I. *Biogeochemical Dynamics at Major River-Coastal Interfaces: Linkages with Global Change* (eds Bianchi, T., Allison, M. & Cai, W.-J.) 353–372 (Cambridge Univ. Press, 2014).
50. Lupker, M., France-Lanord, C., Galy, V., Lavé, J. & Kudrass, H. Increasing chemical weathering in the Himalayan system since the Last Glacial Maximum. *Earth Planet. Sci. Lett.* **365**, 243–252 (2013).
51. Lambeck, K. & Chappell, J. Sea-level change through the last glacial cycle. *Science* **292**, 679–686 (2001).
52. Clark, P. U. et al. The last glacial maximum. *Science* **325**, 710–714 (2009).
53. Collins, J. A. et al. Estimating the hydrogen isotopic composition of past precipitation using leaf-waxes from western Africa. *Quat. Sci. Rev.* **65**, 88–101 (2013).
54. Galy, V., Eglinton, T., France-Lanord, C. & Sylva, S. The provenance of vegetation and environmental signatures encoded in vascular plant biomarkers carried by the Ganges–Brahmaputra rivers. *Earth Planet. Sci. Lett.* **304**, 1–12 (2011).
55. Schwenk, T., Spieß, V., Hübscher, C. & Breitzke, M. Frequent channel avulsions within the active channel–levee system of the middle Bengal Fan—an exceptional channel–levee development derived from Parasound and Hydrosweep data. *Deep Sea Res. Part II Top. Stud. Oceanogr.* **50**, 1023–1045 (2003).

**Acknowledgements** We thank C. Johnson (Woods Hole Oceanographic Institution, WHOI) and D. Montluçon (ETH Zürich) for laboratory support. We thank H. Kudrass for assistance with core sampling. This work was supported by a WHOI Coastal Ocean Institute Postdoctoral Fellowship to C.J.H., and by the National Science Foundation (grant numbers OCE-1333826, OCE-1333387 and OCE-1657771). This is contribution number 3868 of the Virginia Institute of Marine Science.

**Author contributions** C.J.H. and V.V.G. designed the study. C.J.H. conducted laboratory analyses with substantial contributions from V.V.G.; most radiocarbon analyses were conducted by M.U. and N.H. at the Laboratory for Ion Beam Physics (ETH). C.J.H. and V.V.G. drafted the manuscript with contributions from T.I.E.

**Competing interests** The authors declare no competing interests.

### Additional information

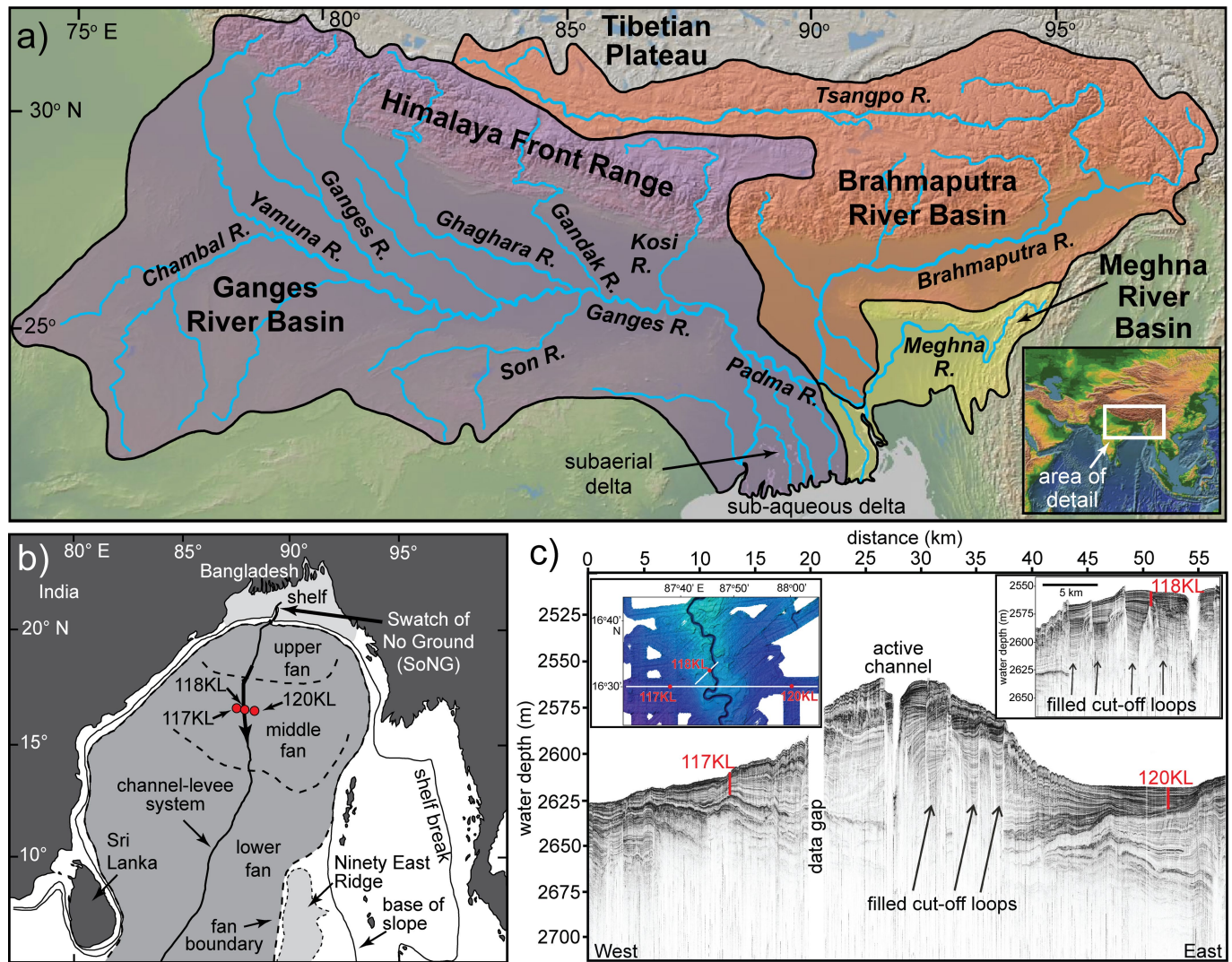
**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41586-020-2233-9>.

**Correspondence and requests for materials** should be addressed to C.J.H.

**Peer review information** Nature thanks Katherine Freeman, Sanjeev Gupta, Yongsong Huang and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

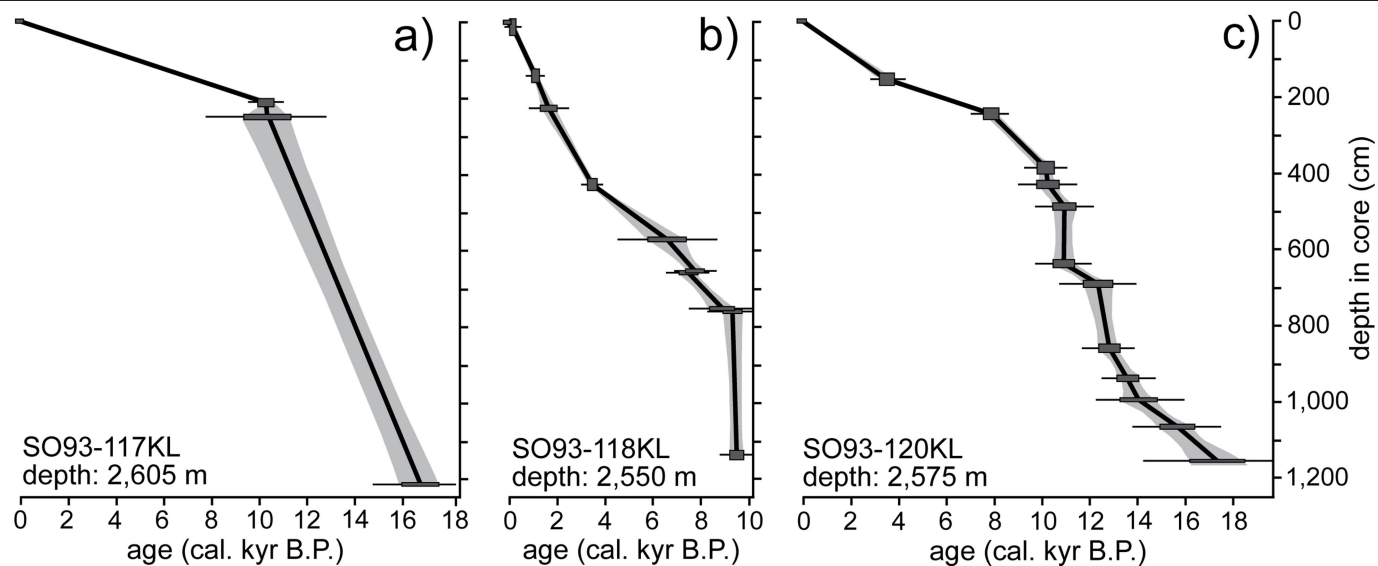
**Reprints and permissions information** is available at <http://www.nature.com/reprints>.





**Extended Data Fig. 1 | Study area and data collection.** **a**, Major features and tributaries of the Ganges–Brahmaputra drainage basin. The background topographic image is from GeoMapApp (<http://www.geomapp.org>). **b**, Morphology of the Bengal Fan. Sediment is dominantly delivered via turbidity currents that travel along the single-channel channel-levee system. Red circles show sediment cores used here. **c**, A Parasound seismic-reflection

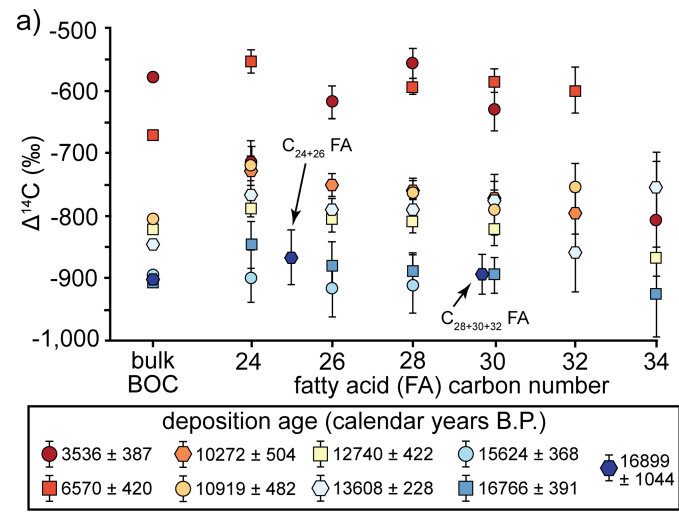
profile crossing cores SO93-117KL and SO93-120KL from west to east. The upper left inset shows locations of the profile and cores with respect to the pathway of the active channel, imaged by multibeam bathymetry. The upper right inset shows Parasound data around core SO93-118KL. Figure parts are modified from ref.<sup>6</sup>, and details of the sediment fan system are described in ref.<sup>55</sup>.



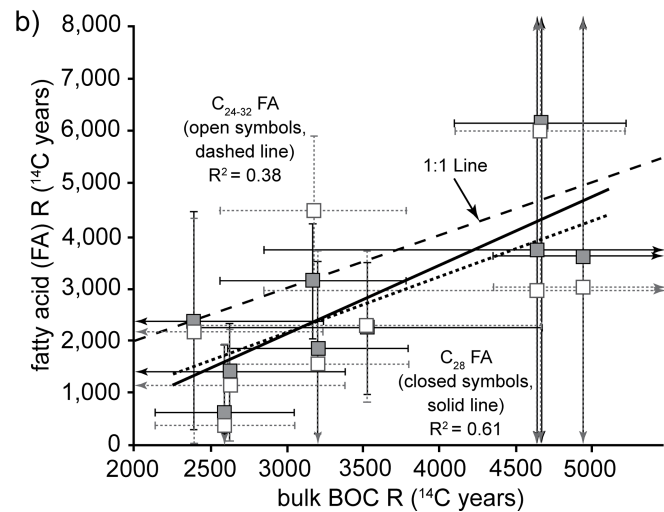
**Extended Data Fig. 2 | Age models for Bengal Fan channel-levee cores.**

**a–c.** Age models for cores SO93-117KL (**a**; number of dated samples,  $n=3$ ), SO93-118KL (**b**;  $n=10$ ) and SO93-120KL (**c**;  $n=12$ , derived through interpolation

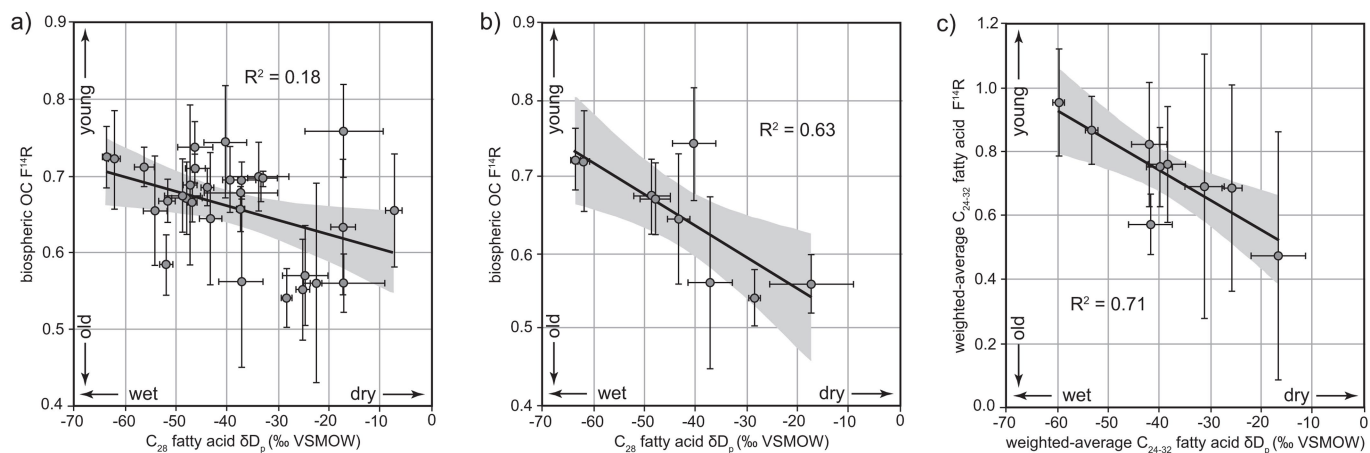
between calibrated  $^{14}\text{C}$  ages (Supplementary Table 1) and extrapolation to core tops and bottoms. Box widths represent sample depth intervals within cores; box heights represent calibrated age errors. Figure updated from ref. <sup>6</sup>.



**Extended Data Fig. 3 | Summary of bulk BOC and biomarker (long-chain fatty acid) analyses.** **a**, Radioisotopic compositions of bulk BOC and individual fatty-acid (FA) homologues ( $n = 47$ ). Vertical error bars indicate propagated measurement uncertainties. **b**, Comparison between reservoir-age offset (as given by the difference between organic-matter and



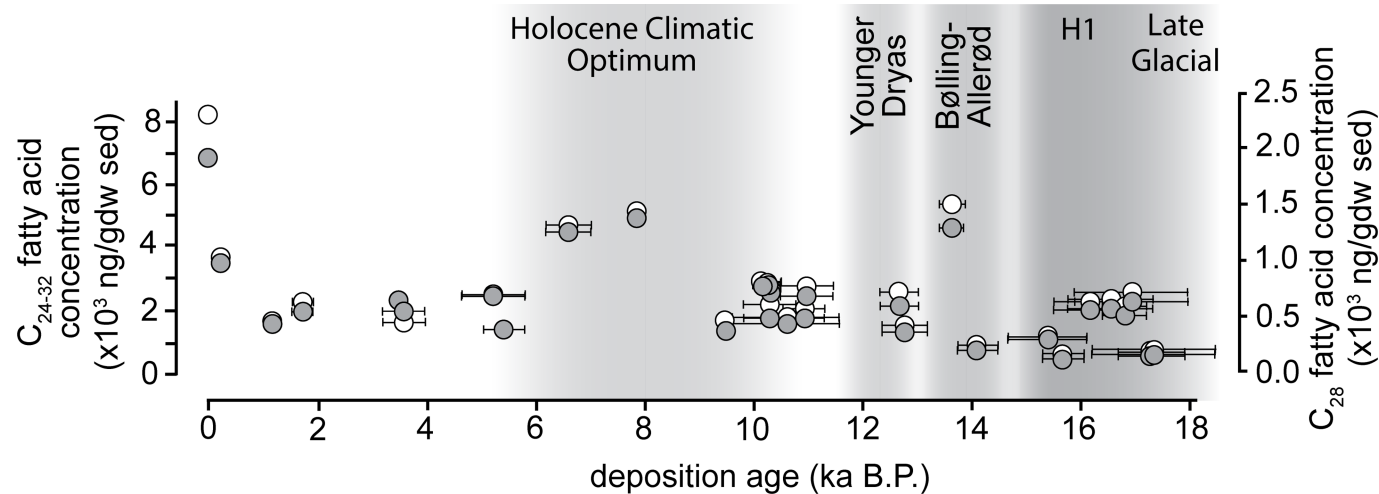
deposition age, in  $^{14}\text{C}$  years) of bulk BOC and  $\text{C}_{28}$  plus  $\text{C}_{24-32}$  fatty acids ( $n = 9$ ). Most values fall below the 1:1 line, reflecting the contribution of excess pre-aged organic matter to the bulk BOC pool. Error bars indicate propagated radiocarbon measurement and instrument-correction uncertainties.



**Extended Data Fig. 4 | Correlation between palaeoclimate proxies with organic-matter age structure.** **a–c**, Graphs show comparisons of post-glacial precipitation  $\delta D_p$  values derived from ice-volume- and vegetation-fractionation-corrected fatty-acid  $\delta D$  values (more-depleted values are indicative of a stronger ISM) and the pre-ageing of organic matter (given as F<sup>14</sup>R values, in dimensionless Fm units; higher values indicate less pre-ageing). Comparisons shown are: **a**, bulk BOC versus  $\delta D_p$  of C<sub>28</sub> fatty acids ( $n = 30$ ) for the comprehensive data set presented

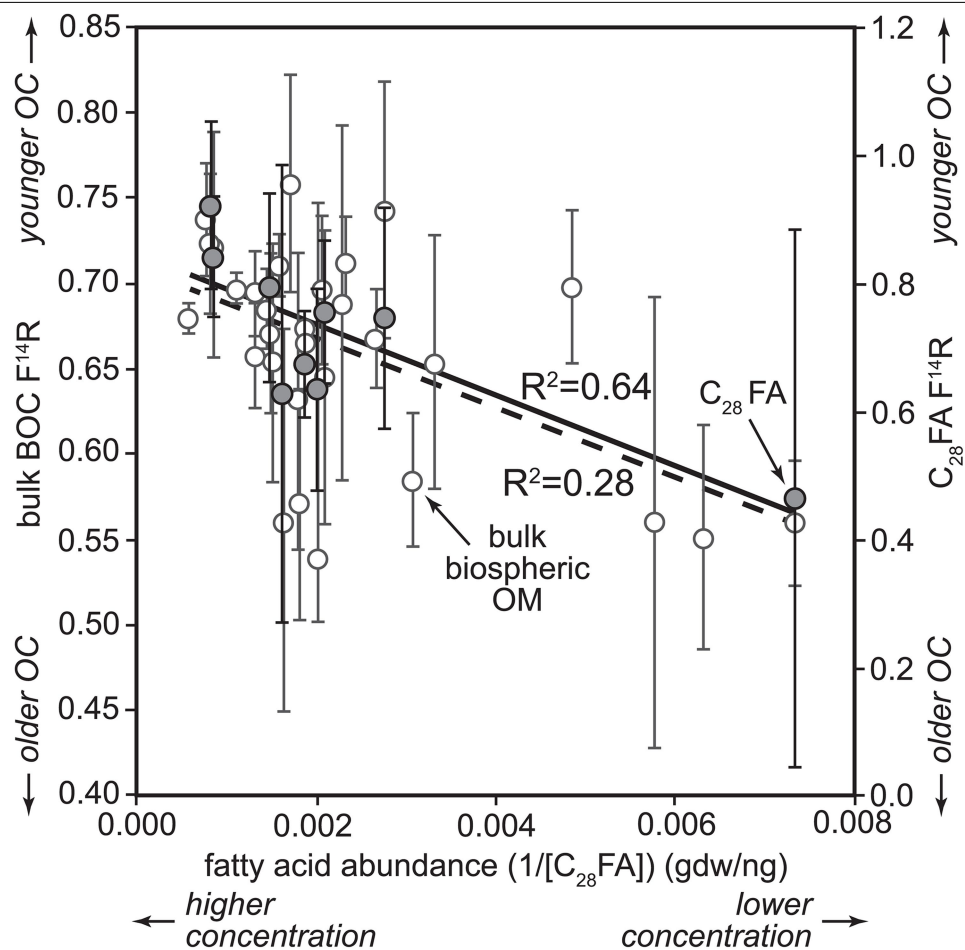
in Supplementary Table 2; **b**, bulk BOC versus  $\delta D_p$  of C<sub>28</sub> fatty acids ( $n = 9$ ) for the subset of samples for which we also have compound-specific (fatty-acid) <sup>14</sup>C data (Supplementary Table 3); and **c**, weighted-average F<sup>14</sup>R values of C<sub>24-32</sub> fatty acid homologues versus weighted-average  $\delta D_p$  values of those same C<sub>24-32</sub> fatty acids ( $n = 9$ ). Vertical error bars indicate propagated radiocarbon measurement and instrument-correction uncertainties; horizontal error bars are propagated multimeasurement standard deviation ( $\delta D$ ) errors (see ref. <sup>6</sup>). OC, organic carbon.





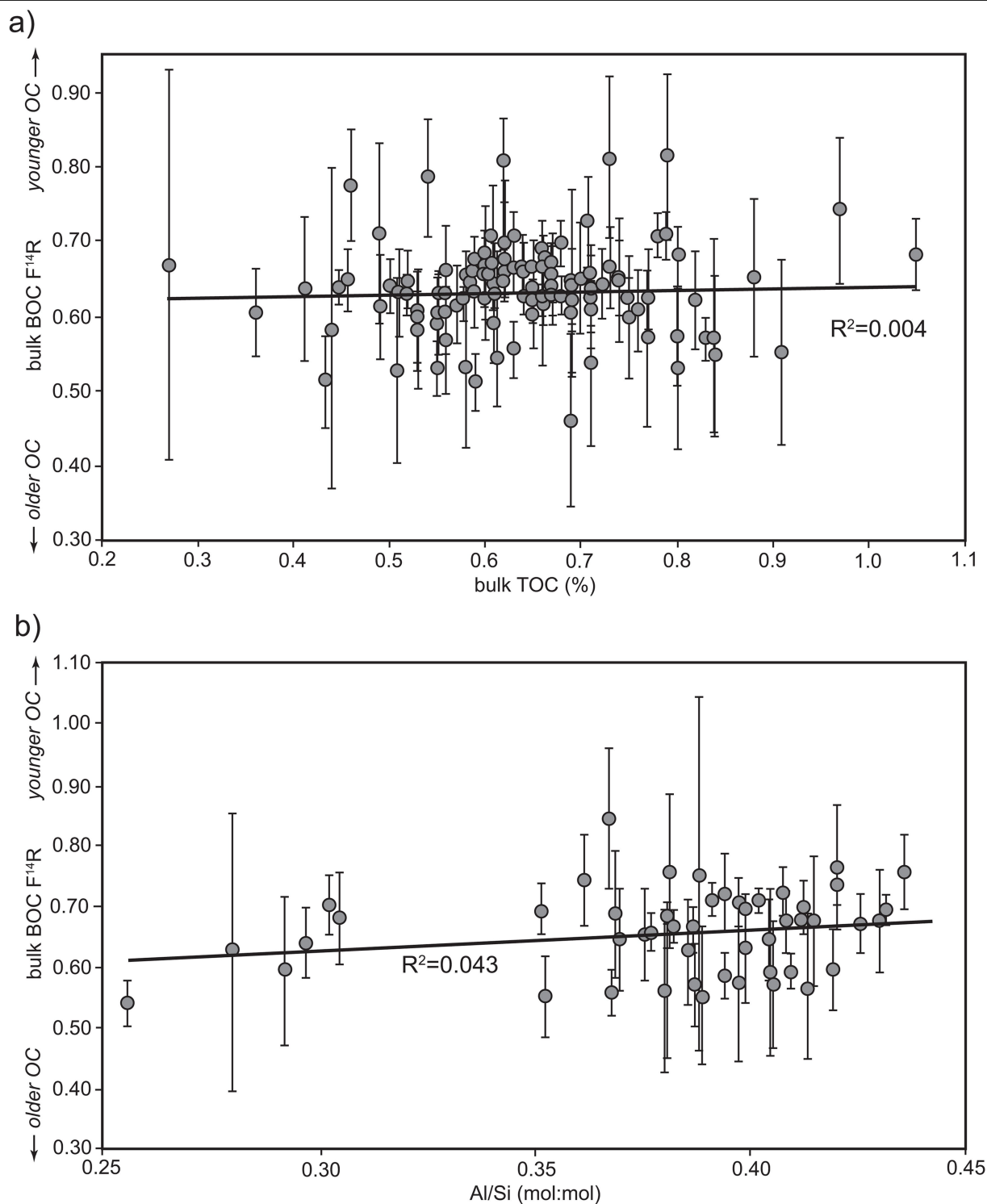
**Extended Data Fig. 5 | Temporal records of fatty acid abundance in Bengal Fan core sediments.** Abundance of  $C_{28}$  (closed circles) and  $C_{24-32}$  (open circles) fatty-acid homologues ( $n=30$ ) in sediments within Bengal Fan channel-levee

cores since the Late Glacial (data given in Supplementary Table 4). Horizontal error bars represent depositional age uncertainties (from core-age models) and are within data points if not shown.



**Extended Data Fig. 6 | Correlation between fatty-acid abundance and organic-matter age structure.**  $C_{28}$  fatty-acid abundances (Supplementary Table 4) are plotted against  $F^{14}R$  values of bulk BOC (open circles, dashed line;

$n = 30$ ) and of  $C_{28}$  fatty acids (closed circles, solid line;  $n = 9$ ). Vertical error bars indicate propagated radiocarbon measurement and instrument-correction uncertainties.



**Extended Data Fig. 7 | Correlation between organic-matter age structure and proxies for sediment and organic-matter composition. a, b,** Bulk BOC  $F^{14}R$  values are plotted against: **a**, bulk sediment TOC values, and **b**, sediment Al/Si values, for all samples used herein for which both data sets exist ( $n = 116$

and  $n = 50$ , respectively). Al/Si values in **b** are from refs. <sup>6,50</sup>. Vertical error bars indicate propagated radiocarbon measurement and instrument-correction uncertainties.

# Tail-propelled aquatic locomotion in a theropod dinosaur

<https://doi.org/10.1038/s41586-020-2190-3>

Received: 17 June 2019

Accepted: 19 February 2020

Published online: 29 April 2020

 Check for updates

Nizar Ibrahim<sup>1</sup>✉, Simone Maganuco<sup>2,3</sup>, Cristiano Dal Sasso<sup>3</sup>, Matteo Fabbri<sup>4</sup>, Marco Auditore<sup>3</sup>, Gabriele Bindellini<sup>3,5</sup>, David M. Martill<sup>6</sup>, Samir Zouhri<sup>7</sup>, Diego A. Mattarelli<sup>3</sup>, David M. Unwin<sup>8</sup>, Jasmina Wiemann<sup>4</sup>, Davide Bonadonna<sup>2</sup>, Ayoub Amame<sup>7</sup>, Juliana Jakubczak<sup>4</sup>, Ulrich Joger<sup>9</sup>, George V. Lauder<sup>10</sup> & Stephanie E. Pierce<sup>10</sup>✉

In recent decades, intensive research on non-avian dinosaurs has strongly suggested that these animals were restricted to terrestrial environments<sup>1</sup>. Historical proposals that some groups, such as sauropods and hadrosaurs, lived in aquatic environments<sup>2,3</sup> were abandoned decades ago<sup>4–6</sup>. It has recently been argued that at least some of the spinosaurids—an unusual group of large-bodied theropods of the Cretaceous era—were semi-aquatic<sup>7,8</sup>, but this idea has been challenged on anatomical, biomechanical and taphonomic grounds, and remains controversial<sup>9–11</sup>. Here we present unambiguous evidence for an aquatic propulsive structure in a dinosaur, the giant theropod *Spinosaurus aegyptiacus*<sup>7,12</sup>. This dinosaur has a tail with an unexpected and unique shape that consists of extremely tall neural spines and elongate chevrons, which forms a large, flexible fin-like organ capable of extensive lateral excursion. Using a robotic flapping apparatus to measure undulatory forces in physical models of different tail shapes, we show that the tail shape of *Spinosaurus* produces greater thrust and efficiency in water than the tail shapes of terrestrial dinosaurs and that these measures of performance are more comparable to those of extant aquatic vertebrates that use vertically expanded tails to generate forward propulsion while swimming. These results are consistent with the suite of adaptations for an aquatic lifestyle and piscivorous diet that have previously been documented for *Spinosaurus*<sup>7,13,14</sup>. Although developed to a lesser degree, aquatic adaptations are also found in other members of the spinosaurid clade<sup>15,16</sup>, which had a near-global distribution and a stratigraphic range of more than 50 million years<sup>14</sup>, pointing to a substantial invasion of aquatic environments by dinosaurs.

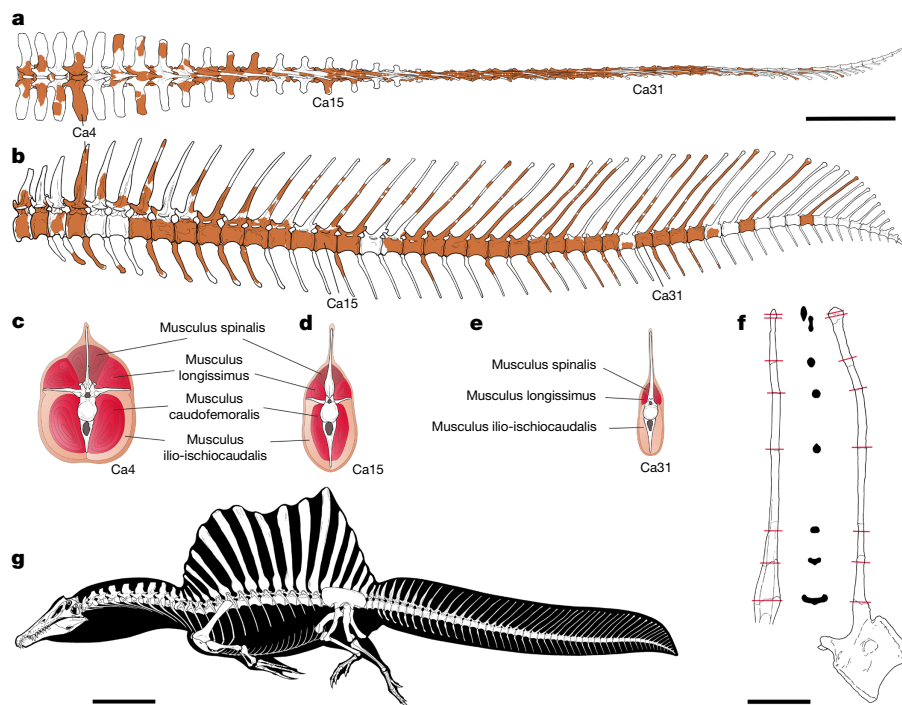
Detailed anatomical and functional studies, combined with abundant trackways, all point to a strictly terrestrial ecology for dinosaurs<sup>1</sup>, with one clade (Maniraptora) taking to the air<sup>17</sup>. Dinosaurs are not currently thought to have invaded aquatic environments, following the abandonment—several decades ago<sup>5,6</sup>—of century-old ideas of semi-aquatic habits in sauropods and hadrosaurs<sup>2,3</sup>. Potential semi-aquatic lifestyles have recently been hypothesized for a small number of dinosaurs<sup>18,19</sup>. However, the only group of dinosaurs for which multiple plausible lines of evidence indicate aquatic adaptations are the spinosaurids, large-bodied theropods interpreted as near-shore waders that fed on fish along the margins of (rather than within) bodies of water<sup>10,15,20</sup>.

A recent study<sup>7</sup> of the largest known spinosaurid, *S. aegyptiacus*, identified a series of adaptations consistent with a semi-aquatic lifestyle, including reduced hindlimbs, wide feet with large flat unguals, long bones with a highly reduced medullary cavity, and a suite of cranial

features (such as retracted nares, interlocking conical teeth and a rostromandibular integumentary sensory system). This interpretation has been challenged on the basis of taphonomy<sup>9</sup>, biomechanical modelling<sup>10</sup> and anatomical concerns<sup>9</sup>. Locomotion in water is a major point of contention<sup>10,11</sup>, because no unambiguous evidence for a plausible mode of propulsion has been presented. Furthermore, our understanding of the anatomy and ecology of this highly derived theropod has been hampered because only one associated *Spinosaurus* skeleton exists, with all other associated remains having been destroyed in World War II<sup>7</sup>. The posterior portion of the skeleton and the caudal vertebral series in particular, which has the potential to shed light on likely adaptations for aquatic locomotion, has until now been poorly understood<sup>12</sup>. Consequently, the tail anatomy and function of *Spinosaurus* has been reconstructed on the basis of highly incomplete remains and potentially spurious comparisons with other similar-sized theropods.

<sup>1</sup>Department of Biology, University of Detroit Mercy, Detroit, MI, USA. <sup>2</sup>Associazione Paleontologica Paleontologica Italiana, Parma, Italy. <sup>3</sup>Sezione di Paleontologia dei Vertebrati, Museo di Storia Naturale di Milano, Milan, Italy. <sup>4</sup>Department of Geology and Geophysics, Yale University, New Haven, CT, USA. <sup>5</sup>Dipartimento di Scienze della Terra 'A. Desio', Università degli Studi di Milano, Milan, Italy. <sup>6</sup>School of the Environment, Geography and Geological Sciences, University of Portsmouth, Portsmouth, UK. <sup>7</sup>Department of Geology, Hassan II University of Casablanca, Casablanca, Morocco. <sup>8</sup>School of Museum Studies, University of Leicester, Leicester, UK. <sup>9</sup>Staatliches Naturhistorisches Museum Braunschweig, Braunschweig, Germany. <sup>10</sup>Museum of Comparative Zoology and Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA, USA. ✉e-mail: [ibrahini@udmercy.edu](mailto:ibrahini@udmercy.edu); [spierce@oeb.harvard.edu](mailto:spierce@oeb.harvard.edu)





**Fig. 1 | Reconstructed skeleton and caudal series of FSAC-KK 11888.**

**a, b**, Caudal series (preserved parts shown in colour) in dorsal view (**a**) and left lateral view (**b**). **c–e**, Reconstructed sequential cross-sections through the tail show proximal-to-distal changes in the arrangement of major muscles.

**f**, Sequential cross-sections through the neural spine of caudal vertebra 23 (Ca23) to show apicobasal changes. **g**, Skeletal reconstruction. Scale bars, 50 cm (**a–e**), 10 cm (**f**), 1 m (**g**).

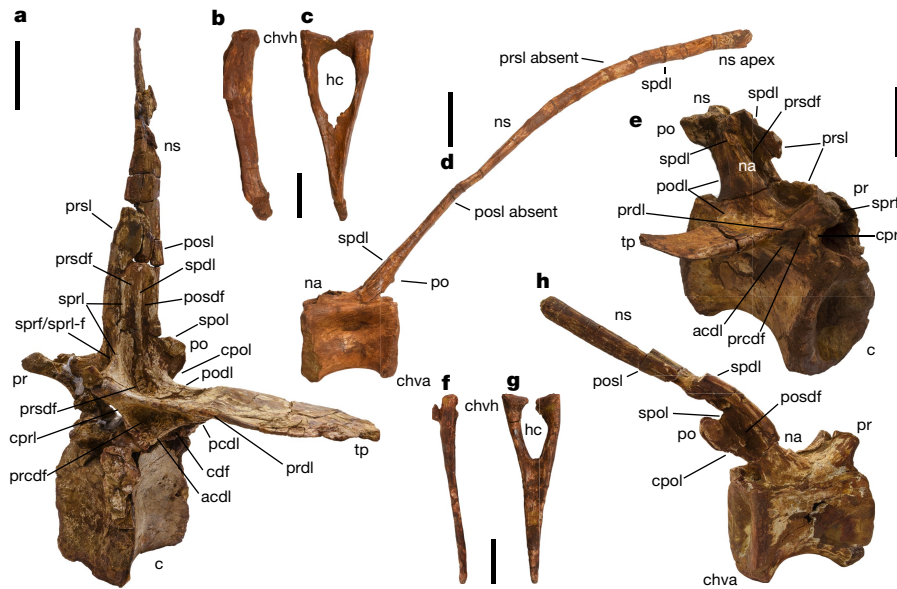
Here we describe a nearly complete and partially articulated tail of a subadult individual of *S. aegyptiacus* (accession code Faculté des Sciences de Casablanca University (FSAC)-KK 11888), from the Cretaceous Kem Kem beds of south-eastern Morocco (Figs. 1, 2, Extended Data Figs. 1–4, Supplementary Information section 1, Supplementary Video 1). The skeleton represents, to our knowledge, the most complete dinosaur known from the Kem Kem beds<sup>21,22</sup> and the most complete skeleton of a Cretaceous theropod known from mainland Africa (Supplementary Information section 2). As we show here, the tail forms part of the neotype of *S. aegyptiacus*<sup>7</sup> and was found in direct juxtaposition to the remainder of the skeleton (Extended Data Fig. 3). The newly recovered material confirms the previous conclusion<sup>7</sup> that a single subadult individual is preserved at the site; over 90% of the new material was recovered during field excavations in late 2018, and then digitally recorded (Extended Data Figs. 1–4, Supplementary Information sections 2–5). Several elements conform closely to drawings of the *Spinosaurus* fossils that were destroyed in World War II (Extended Data Fig. 6).

More than 30 near-sequential caudal vertebrae (located within caudal positions 1–41) of FSAC-KK 11888 are preserved, and represent approximately 80% of the original tail length (Extended Data Figs. 3, 4, Extended Data Tables 1, 2). Both proximal and distal elements of the tail are complete and preserved in three dimensions, indicating minimal taphonomic distortion (Fig. 2, Supplementary Video 2). At the level of the caudal transition point<sup>1</sup>, the centra become proportionally more elongate. In addition, the prezygapophyses no longer overhang the preceding centrum and show a marked decrease in size compared to those of many theropod dinosaurs<sup>1</sup>. The postzygapophyses also decrease in size (Fig. 2), leading to a reduced contact with the prezygapophyses, and are absent in the distalmost caudal vertebrae. This again is different from the condition seen in most theropods, in which zygapophyses become more elongate and more prominent towards the tail tip<sup>1</sup>, restricting flexibility in more distal intervertebral joints.

The neural arches are distinctive elements of the *Spinosaurus* tail. A notably complex array of vertebral laminae and fossae is present in

the proximal caudal vertebrae, and partly persists in mid-caudal neural arches. The morphology of the neural spines shows considerable variation along the sequence (Figs. 1, 2, Extended Data Table 1): the spines of the proximal caudal vertebrae are about three times taller than their centra and are cross-shaped in cross-section from their base to mid-height; in mid-caudal vertebrae, the spines become much longer; and in the small distal caudal vertebrae, the length of the neural spines reaches well over seven times the height of the centrum, in contrast to the condition suggested in a previous study<sup>11</sup>. The neural spines of mid-distal caudal vertebrae of *Spinosaurus* have a cross-section that is unique among theropods: they are proximodistally—rather than transversely—flattened. This is owing to the hyper-development of the spinodiapophyseal laminae and the loss of pre- and postspinal laminae. The chevrons also differ from those of other theropods. The morphology of the chevrons in *Spinosaurus* varies little throughout the caudal series, except for a slight gradual reduction of the haemal canal: distal chevrons are as elongate as the proximal chevrons (Extended Data Table 2) but become slender, paralleling the gradual decrease in the size of the centra. Taken together, the elongate neural and haemal arches result in a tail shape that is markedly vertically expanded and has an extensive lateral surface area (Fig. 1, Extended Data Figs. 3, 4).

The skeletal anatomy of *Spinosaurus* represents a major departure from that of other theropods—including from that of other members of the Tetanurae clade (which comprises crown group birds and all other stem theropods more closely related to birds than to *Ceratosaurus*<sup>1</sup>). One feature of the Tetanurae is a stiffened tail in which the degree of overlap in articulation between pre- and postzygapophyses increases along the caudal series, greatly diminishing the range of motion between individual vertebrae<sup>1</sup>. This trend toward reduced mobility is emphasized in paravians, with the appearance of ossified ligaments and/or reduction and fusion of the caudal vertebrae into a pygostyle<sup>17</sup>. By contrast, in *Spinosaurus* the pre- and postzygapophyses are much further reduced than in other tetanurans and—in the middle



**Fig. 2 | Selected caudal vertebrae and chevrons of FSAC-KK 11888.**

**a**, Proximal caudal vertebra (Ca4) in left proximolateral view. **b, c**, Proximal chevron (Chv7) in left lateral (**b**) and proximal (**c**) view. **d**, Distal caudal vertebra (Ca31) in left lateral view. **e**, Mid-caudal vertebra (Ca12) in right proximolateral view. **f, g**, Distal chevron (Chv27) in left lateral (**f**) and proximal (**g**) views. **h**, Mid-caudal vertebra (Ca16) in right distolateral view. acdl, anterior centrodiapophyseal lamina; c, centrum; ca, caudal vertebra; cdf, centrodiapophyseal fossa; chva, chevron articulation; chvh, chevron head; cpol, centropostzygapophyseal lamina; cprl, centroprezygapophyseal lamina; hc, haemal canal; na, neural arch; ns, neural spine; pcdl, posterior

centrodiapophyseal lamina; po, postzygapophysis; pcdfl, postzygapophyseal centrodiapophyseal fossa; podl, postzygodiapophyseal lamina; posdf, postzygapophyseal spinodiapophyseal fossa; posl, postspinal lamina; pr, prezygapophysis; prcdf, prezygapophyseal centrodiapophyseal fossa; prdl, prezygodiapophyseal lamina; prsdf, prezygapophyseal spinodiapophyseal fossa; prsl, prespinal lamina; spdli, spinodiapophyseal lamina; spol, spinopostzygapophyseal lamina; spof, spinopostzygapophyseal fossa; sprl, spinoprezygapophyseal lamina; sprl-f, spinoprezygapophyseal lamina fossa; tp, transverse process. Scale bars, 10 cm (**a**), 5 cm (**b–h**).

and distal portions of the tail—not only do not overlap, but almost disappear (Fig. 2); this allows the caudal region considerable flexibility, especially with regard to lateral movements.

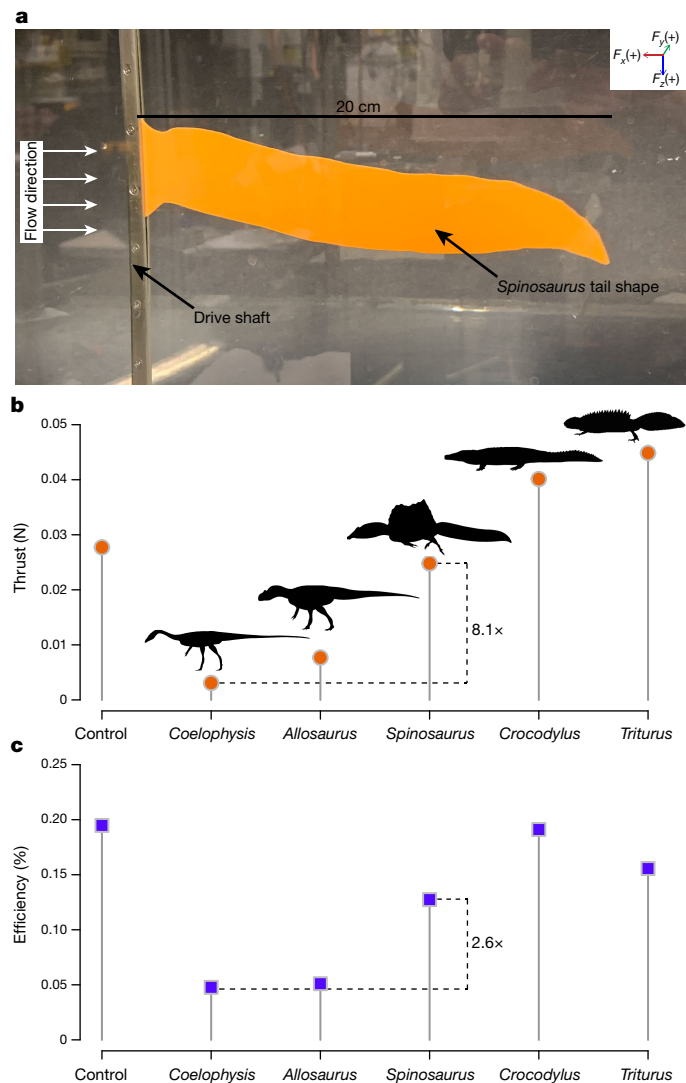
We hypothesized that the highly specialized morphology of the *Spinosaurus* tail allowed it to function as a propulsive structure for aquatic locomotion. To test this idea, we evaluated the swimming potential of the *Spinosaurus* tail shape by comparing it to the tails of two terrestrial theropods (*Coelophysis bauri* and *Allosaurus fragilis*), two semi-aquatic tetrapods (the crocodile *Crocodylus niloticus* and the crested newt *Triturus dobrogicus*) and a rectangular control. Two-dimensional tail shapes were cut from 0.93-mm-thick plastic of flexural stiffness  $5.8 \times 10^{-5} \text{ Nm}^2$ . The plastic tails were attached to a robotic controller and actuated in a water flume to provide tail-tip amplitudes that were approximately 40% of tail length during swimming at 0.5 tail lengths per second. This swimming speed and amplitude of motion is similar to that of slow aquatic locomotion in modern tetrapods<sup>23–25</sup>. We measured swimming performance by quantifying the mean thrust and efficiency using a six-axis force–torque sensor attached to the shaft that drove each tail shape<sup>26</sup> (Fig. 3, Methods, Supplementary Fig. 4, Supplementary Videos 3–5).

Our experimental results show that the *Spinosaurus* tail shape was capable of generating more than 8 times the thrust of the tail shapes of other theropods, and achieved 2.6 times the efficiency (Fig. 3, Supplementary Data 1). The greatest thrust was achieved by the tail shape of the crested newt (1.8 times that of *Spinosaurus* and 14.8 times that of *Coelophysis*), but the crocodile tail shape achieved greater propulsive efficiency (1.5 times that of *Spinosaurus* and 4.0 times that of *Coelophysis*), comparable to the rectangular control (Fig. 3). The lower efficiency recovered in this experiment for *Spinosaurus* (compared to the control with the same surface area) and the crested newt indicates an effect of tail shape on performance. Overall, the vertically expanded tail shape of *Spinosaurus* imparts a substantial positive benefit to aquatic propulsion relative to the long and narrow tails of terrestrial theropods, supporting

the inference that *Spinosaurus* used tail-propelled swimming. This tail morphology may have also increased the lateral stability of the body in the water, reducing the tendency to roll while floating<sup>10</sup>.

Contrary to recent suggestions<sup>10</sup> that *Spinosaurus* was confined to wading and the apprehension of prey from around the edges of bodies of water, the morphology and function of its tail—along with its other adaptations for life in water<sup>7</sup>—point to *Spinosaurus* having been an active and highly specialized aquatic predator that pursued and caught its prey in the water column (Extended Data Fig. 7). The skeletal remains of *Spinosaurus* (Supplementary Information) from the Kem Kem beds—composed of sediments deposited in a major fluvio-deltaic system<sup>7</sup> that have yielded a diverse vertebrate assemblage<sup>27</sup>—provide further insights into the ecology of this dinosaur. The composition of the ecosystem represented by the Kem Kem assemblage is highly atypical, containing a rich freshwater fauna dominated by fishes (including lungfish and large-to-very-large sawfish and coelacanths<sup>27</sup>), a diverse range of crocodyliforms<sup>28</sup> and several giant predatory dinosaurs<sup>7,22</sup>. The seemingly anomalous occurrence in the same deposits of several large-bodied predators but few terrestrial herbivores is partially explained by the largely aquatic and probably piscivorous lifestyle of *Spinosaurus*, which considerably expands the morphological and ecological disparity of Kem Kem tetrapods<sup>7,29</sup>. At the same time, competition with several co-occurring large aquatic predators<sup>28</sup> may have driven the evolution of giant size in *Spinosaurus*.

Although the unique postcranial adaptations of *Spinosaurus* point towards an entirely novel mode of locomotion within Dinosauria, other spinosaurids share a wide range of derived anatomical features that are consistent with a partially aquatic, piscivorous mode of life<sup>7,8,11,14,30</sup>. The exact extent to which an aquatic lifestyle was adopted by these taxa, and how this varied across Spinosauridae, remains to be established. However, the near-global distribution of spinosaurids (which have now been reported from Europe, Asia, Africa and South America<sup>30</sup>) and their



**Fig. 3 | Comparative tail swimming performance.** **a**, *Spinosaurus* plastic tail shape attached to a robotic drive shaft in a water flume with water flowing at  $10 \text{ cm s}^{-1}$ . With reference to the tail, positive  $x$  forces ( $F_x$ ) are generated proximally (or upstream), positive  $y$  forces ( $F_y$ ) in the right lateral direction, and positive  $z$  forces ( $F_z$ ) in the ventral direction. **b**, **c**, Mean thrust (**b**) and mean efficiency (**c**) generated by tail shapes during robotically controlled swimming. All tails are scaled to 20-cm length (Supplementary Fig. 4). The control tail was rectangular in shape with the same surface area as the scaled *Spinosaurus* tail ( $63 \text{ cm}^2$ ). See Methods for experimental setup. Raw thrust and efficiency data, including mean and s.e.m., are provided in Supplementary Data 1. Swimming motions of the *Spinosaurus* tail are visualized in Supplementary Videos 3–5.

substantial temporal range (first appearing, based on phylogenetic inference, in the Mid–or possibly even Early–Jurassic epoch and with a fossil record that spans more than 50 million years (from the Late Jurassic to the early Late Cretaceous epoch)<sup>14</sup>) point to a persistent and widespread invasion of aquatic habitats by dinosaurs.

## Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information,

acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-020-2190-3>.

- Weishampel, D. B., Dodson, P. & Osmólska, H. *The Dinosauria* 2nd edn (Univ. of California Press, Berkeley, 2004).
- Owen, R. A description of a portion of the skeleton of the *Cetiosaurus*, a gigantic extinct saurian reptile occurring in the oolitic formations of different portions of England. *Proc. Geol. Soc. Lond.* **3**, 457–462 (1841).
- Cope, E. On the characters of the skull in the Hadrosauridae. *Proc. Acad. Nat. Sci. Philadelphia* **35**, 97–107 (1883).
- Kermack, K. A. A note on the habits of sauropods. *Ann. Mag. Nat. Hist.* **4**, 830–832 (1951).
- Bakker, R. T. Ecology of the brontosaurs. *Nature* **229**, 172–174 (1971).
- Alexander, R. M. Mechanics of posture and gait of some large dinosaurs. *Zool. J. Linn. Soc.* **83**, 1–25 (1985).
- Ibrahim, N. et al. Semiaquatic adaptations in a giant predatory dinosaur. *Science* **345**, 1613–1616 (2014).
- Aureliano, T. et al. Semi-aquatic adaptations in a spinosaur from the Lower Cretaceous of Brazil. *Cretac. Res.* **90**, 283–295 (2018).
- Evers, S. W., Rauhut, O. W. M., Milner, A. C., McFeeters, B. & Allain, R. A reappraisal of the morphology and systematic position of the theropod dinosaur *Sigilmassasaurus* from the “middle” Cretaceous of Morocco. *PeerJ* **3**, e1323 (2015).
- Henderson, D. M. A buoyancy, balance and stability challenge to the hypothesis of a semi-aquatic *Spinosaurus* Stromer, 1915 (Dinosauria: Theropoda). *PeerJ* **6**, e5409 (2018).
- Hone, D. W. E. & Holtz, T. R. Jr. Comment on: Aquatic adaptation in the skull of carnivorous dinosaurs (Theropoda: Spinosauridae) and the evolution of aquatic habits in spinosaurids. 93: 275–284. *Cretac. Res.* <https://doi.org/10.1016/j.cretres.2019.05.010> (2019).
- Stromer, E. Ergebnisse der Forschungsreisen Prof. E. Stromers in den Wüsten Ägyptens. II. Wirbeltier-Reste der Baharije - Stufe (unterstes Cenoman). 3. Das Original des Theropoden *Spinosaurus aegyptiacus* nov. gen., nov. spec. *Abh. Kgl. Bayer. Akad. Wiss. Math. Phys. Kl. München* **28**, 1–28 (1915).
- Vullo, R. et al. Convergent evolution of jaws between spinosaurid dinosaurs and Pike Conger Eels. *Acta Palaeontol. Pol.* **61**, 825–829 (2016).
- Arden, T. M. S., Klein, C. G., Zouhri, S. & Longrich, N. R. Aquatic adaptation in the skull of carnivorous dinosaurs (Theropoda: Spinosauridae) and the evolution of aquatic habits in *Spinosaurus*. *Cretac. Res.* **93**, 275–284 (2019).
- Charig, A. J. & Milner, A. C. *Baryonyx walkeri*, a fish-eating dinosaur from the Wealden of Surrey. *Bull. Nat. Hist. Mus. Geol.* **53**, 11–70 (1997).
- Sues, H. D., Frey, E., Martill, D. M. & Scott, D. M. *Irritator challengeri*, a spinosaurid (Dinosauria: Theropoda) from the Lower Cretaceous of Brazil. *J. Vert. Pal.* **22**, 535–547 (2002).
- Witmer, L. M. in *Mesozoic Birds: Above the Heads of Dinosaurs* (eds Chiappe, L. M. & Witmer, L. M.) 3–30 (Univ. California Press, 2002).
- Tereschenko, V. Adaptive features of protoceratopsids (Ornithischia, Neoceratopsia). *Paleontol. J.* **42**, 273–286 (2008).
- Cau, A. et al. Synchrotron scanning reveals amphibious ecomorphology in a new clade of bird-like dinosaurs. *Nature* **552**, 395–399 (2017).
- Sereno, P. C. et al. A long-snouted predatory dinosaur from Africa and the evolution of spinosaurids. *Science* **282**, 1298–1302 (1998).
- Lavocat, R. Sur les dinosauriens du Continental Intercalaire des Kem-Kem de la Daour. In *Comptes rendus 19e Congrès Géologique International, Alger, 1952* 65–68 (Académie des Sciences de Paris, 1954).
- Sereno, P. C. et al. Predatory dinosaurs from the Sahara and Late Cretaceous faunal differentiation. *Science* **272**, 986–991 (1996).
- D’Aout, K. & Aerts, P. Kinematic and swimming efficiency of steady swimming in adult axolotls (*Ambystoma mexicanum*). *J. Exp. Biol.* **200**, 1863–1871 (1997).
- Fish, F. Kinematics of undulatory swimming in the American alligator. *Copeia* **1984**, 839–843 (1984).
- Frolich, L. M. & Biewener, A. A. Kinematic and electromyographic analysis of the functional role of the body axis during terrestrial and aquatic locomotion in the salamander *Ambystoma tigrinum*. *J. Exp. Biol.* **162**, 107–130 (1992).
- Lauder, G. V., Flammang, B. & Alben, S. Passive robotic models of propulsion by the bodies and caudal fins of fish. *Integr. Comp. Biol.* **52**, 576–587 (2012).
- Cavin, L. et al. Vertebrate assemblages from the early Late Cretaceous of southeastern Morocco: an overview. *J. Afr. Earth Sci.* **57**, 391–412 (2010).
- Meunier, L. M. V. & Larsson, H. C. E. Revision and phylogenetic affinities of *Elosuchus* (Crocodyliformes). *Zool. J. Linn. Soc.* **179**, 169–200 (2017).
- Amiot, R. et al. Oxygen and carbon isotope compositions of middle Cretaceous vertebrates from North Africa and Brazil: ecological and environmental significance. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **297**, 439–451 (2010).
- Hone, D. W. E. & Holtz, T. R. A century of spinosaurs. A review and revision of the Spinosauridae with comments on their ecology. *Acta Geol. Sin.* **91**, 1120–1132 (2017).

**Publisher’s note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2020



## Methods

No statistical methods were used to predetermine sample size. The experiments were not randomized and investigators were not blinded to allocation during experiments and outcome assessment.

### Excavation and reconstruction

The Cretaceous Kem Kem beds of Morocco crop out along an extensive escarpment, often located near the Moroccan–Algerian border<sup>7</sup>. After the accidental discovery and partial excavation by a local collector in 2008, part of a single skeleton (FSAC-KK 11888)—subsequently deposited at the Faculté des Sciences of Casablanca University—was recovered, published and designated as the neotype of *S. aegyptiacus*<sup>7</sup>. A multi-institutional collaborative project in the years 2015–2019, led by N.I., resulted in four joint expeditions to the neotype site. Detailed and careful exploration of the debris around the site, as well as a systematic and extended excavation of the unexposed portion of the fossiliferous layer of the Zrigat hill, led to the recovery of many additional elements of the neotype skeleton (Extended Data Figs. 1–5). A detailed description of the newly recovered material, as well as the geological context, is included in the Supplementary Information. The Supplementary Information also includes details of a full-body flesh reconstruction of *Spinosaurus* based on FSAC-KK 11888, as well as estimates of whole-body mass, segment masses, segment centres of mass and whole-body centre of mass (Supplementary Data 2). The position of the centre of mass in comparison to that in previously published analyses<sup>7,10</sup> can be found in Extended Data Fig. 8.

### Osteohistological analysis

The aim of the osteohistological analysis was to determine whether the remains that were assigned to FSAC-KK 11888 belong to a single individual, rather than a chimeric association of juvenile and adult individuals preserved in the same location and at the same horizon. The analysis was based on five skeletal elements. The primary assumption is that, should histological details suggest that all five elements represent the same ontogenetic stage, then the remains are more likely to represent one individual than multiple individuals. By contrast, should these elements exhibit two or more distinct ontogenetic stages, this would point to the presence of multiple individuals of one taxon (or perhaps several taxa), all fortuitously preserved at a single location during a single depositional event<sup>31–33</sup>.

The following elements were sectioned: the right femur; the left fibula; one rib; and two neural spines. All specimens were sectioned before preparation to ensure that no outer layers of the compact cortex were accidentally removed. In the case of the neural spines, the apical portion was sectioned.

Thin sectioning followed standard protocols<sup>34</sup>. The thin sections have a thickness of 50–70 µm, and were analysed with a petrographic microscope (Leica DM 2500 P). Digital images were captured using a ProgRes Cfsan camera. Only continuous lines were counted as lines of arrested growth. Annuli were interpreted as a single year, following a previous publication<sup>35</sup>. Retrocalculation, following a previously published method<sup>36</sup>, was applied to determine the likely number of missing lines of arrested growth, eroded through remodelling of the bone. In the case of the neural spines, only the width of the innermost zone was used to retrocalculate the missing lines of arrested growth, because the shape of the section could not be approximated to a circular outline. The calculation of the major and minor axes used for the retrocalculation was performed in ImageJ<sup>37</sup>. Results of the histological analysis are included in the Supplementary Information.

### Experimental testing of tail-shape swimming performance

To test the aquatic locomotor potential of the newly reconstructed *S. aegyptiacus* tail, we determined the swimming performance of its tail shape using a robotic controller developed for studies of propulsive

hydrodynamics<sup>38–42</sup>. The swimming performance of the *Spinosaurus* tail shape was compared to the performance of five other tail shapes from the following species: the small-bodied terrestrial theropod *C. bauri*, the large-bodied terrestrial theropod *A. fragilis*, the semi-aquatic crocodile *C. niloticus*, the semi-aquatic crested newt *T. dobrogicus* and a rectangular control tail that was scaled to the same surface area as the *Spinosaurus* tail. Tail shapes (Supplementary Fig. 4) were all scaled to 20 cm proximodistal length, manufactured from 0.93-mm-thick plastic of flexural stiffness  $5.8 \times 10^{-5} \text{ Nm}^2$  and cut using an Epilog Zing 24 laser cutter.

The plastic tails were attached to a robotic controller that allowed us to impose specific motion programmes on the rigid shaft to which each tail was affixed (Fig. 3, Supplementary Videos 3–5). This shaft was moved in both heave (side-to-side) motion, as well as in pitch (angular rotation), to achieve undulatory tail motions. The imposed motion programme was 1-Hz frequency,  $\pm 1\text{-cm}$  heave and  $\pm 25^\circ$  pitch, which resulted in the tail tip undergoing peak-to-peak lateral excursions of approximately 40% of the proximodistal length, comparable to that exhibited by swimming axolotls and alligators<sup>23–25</sup>.

The shaft supporting each simulated tail at the leading edge was attached to an ATI (Apex) Nano-17 six-axis force–torque sensor located just above the water surface. Testing occurred in a recirculating water flume, and a free-stream flow of  $0.5 \text{ l (10 cm s}^{-1})$  was imposed for all tests. Custom LabVIEW programs (National Instruments) were used to control flapping frequency, flow speed, heave and pitch. A custom LabVIEW program also was used to acquire data from the ATI transducer at a sampling rate of 1,000 Hz. Each tail shape was tested  $n = 5$  times (except for the *Spinosaurus* tail, which was tested  $n = 5$  times on 2 different days for a total of  $n = 10$  tests). Output data can be found in Supplementary Data 1.

Thrust and efficiency for each tail shape were calculated using standard fluid dynamic equations as in previous research<sup>43,44</sup>. Mean thrust force ( $F_x$ ) is calculated directly from transducer output from the  $F_x$  channel, and we accounted for transducer rotation resulting from the pitch motion to provide the force component directed upstream (positive thrust). Propulsive efficiency is calculated as the ratio of the thrust coefficient ( $C_T = 2F_x/\rho U^2 cs$ ) to the power coefficient ( $C_p = 2\bar{P}/\rho U^3 cs$ ), in which  $\rho$  is the fluid density,  $U$  is the swimming velocity,  $c$  is the foil chord and  $s$  is the tail span. Effectively, this metric assesses the extent to which input power is translated into thrust.

### Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

### Data availability

The authors declare that all data supporting the findings of this study are available in the paper and its Supplementary Information. Three-dimensional data are available on SketchFab: flesh model at <https://sketchfab.com/3d-models/07b2b6bf4c464c09bd30daa629f266ff>; scanned caudal vertebrae and chevrons at <https://sketchfab.com/3d-models/chv-7-ca70592e5d07408980220d639bc1456f> (Chv7), <https://sketchfab.com/3d-models/chv-24-d917f541a7934492aa0be7c5b97ad40> (Chv24), <https://sketchfab.com/3d-models/ca-4-56e19d32f53043369ba23d5283279eef> (Ca4), <https://sketchfab.com/3d-models/ca-7-c2f551b61e294138954c4f2224bd3353> (Ca7), <https://sketchfab.com/3d-models/ca-12-bb2f30fec9064645b2ff99be30f1ac92> (Ca12), <https://sketchfab.com/3d-models/ca-16-3cd5f6713e1f43f5bf466471416f06c8> (Ca16), <https://sketchfab.com/3d-models/ca-23-f34b71eaf4e54cb29a89a0f50730e70a> (Ca23), <https://sketchfab.com/3d-models/ca-24-4e21ceca81c8403484b9b7e3da7bd0d6> (Ca24), <https://sketchfab.com/3d-models/ca-31-6d0748a851994d6d86b84803743b75a4> (Ca31) and <https://sketchfab.com/3d-models/ca-41-34e5e6415f7d4fce-a335e7120b9fe9b4> (Ca41).



31. Ryan, M. J., Russell, A. P., Eberth, D. A. & Currie, P. J. The taphonomy of a *Centrosaurus* (Ornithischia: Ceratopsidae) bone bed from the Dinosaur Park Formation (Upper Campanian), Alberta, Canada, with comments on cranial ontogeny. *Palaios* **16**, 482–506 (2001).
32. Erickson, G. M., Currie, P. J., Inouye, B. D. & Winn, A. A. Tyrannosaur life tables: an example of nonavian dinosaur population biology. *Science* **313**, 213–217 (2006).
33. Bertozzo, F., Dalla Vecchia, F. M. & Fabbri, M. The Venice specimen of *Ouranosaurus nigeriensis* (Dinosauria, Ornithopoda). *PeerJ* **5**, e3403 (2017).
34. Chinsamy, A. & Raath, M. A. Preparation of fossil bone for histological examination. *Palaeontol. Afr.* **29**, 39–44 (1992).
35. Lee, A. H. & O'Connor, P. M. Bone histology confirms determinate growth and small body size in the noasaurid theropod *Masiakasaurus knopfleri*. *J. Vertebr. Paleontol.* **33**, 865–876 (2013).
36. Horner, J. R. & Padian, K. Age and growth dynamics of *Tyrannosaurus rex*. *Proc. R. Soc. Lond. B* **271**, 1875–1880 (2004).
37. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **9**, 671–675 (2012).
38. Lauder, G. V., Anderson, E. J., Tangorra, J. & Madden, P. G. A. Fish biorobotics: kinematics and hydrodynamics of self-propulsion. *J. Exp. Biol.* **210**, 2767–2780 (2007).
39. Lauder, G. V. et al. Robotic models for studying undulatory locomotion in fishes. *Mar. Technol. Soc. J.* **45**, 41–55 (2011).
40. Quinn, D. B., Lauder, G. V. & Smits, A. J. Maximizing the efficiency of a flexible propulsor using experimental optimization. *J. Fluid Mech.* **767**, 430–448 (2015).
41. Rosic, M. N., Thornycroft, P. J. M., Feilich, K. L., Lucas, K. N. & Lauder, G. V. Performance variation due to stiffness in a tuna-inspired flexible foil model. *Bioinspir. Biomim.* **12**, 016011 (2017).
42. Saadat, M. et al. On the rules for aquatic locomotion. *Phys. Rev. Fluids* **2**, 083102 (2017).
43. Read, D. A., Hover, F. S. & Triantafyllou, M. S. Forces on oscillating foils for propulsion and maneuvering. *J. Fluids Structures* **17**, 163–183 (2003).
44. Shelton, R. M., Thornycroft, P. J. & Lauder, G. V. Undulatory locomotion of flexible foils as biomimetic models for understanding fish propulsion. *J. Exp. Biol.* **217**, 2110–2120 (2014).

**Acknowledgements** We thank M. Azroal, H. Azroal, M. Fouadassi and all other expedition members from the 2015, 2018 and 2019 seasons for assistance in the field; A. A. Ha for help in preparing the fossils; the Moroccan Ministry of Mines, Energy and Sustainable Development for providing fieldwork permits; F. Manucci for helpful discussions about the flesh reconstruction of *Spinosaurus*; and P. Fahn-Lai for coding assistance. This research was supported by a National Geographic Society grant to N.I. (CP-143R-170), a National Geographic Emerging Explorer Grant to N.I., contributions from the Board of Advisors of the University of Detroit Mercy to N.I., a Jurassic Foundation grant to M.F., a Paleontological Society grant to M.F., an Explorers Club grant to M.F., as well as financial support from the Lokschnuppen Rosenheim, the Museo di Storia Naturale di Milano, J. Pfauntsch and A. Lania.

**Author contributions** N.I. led the expeditions and the project. N.I., S.M., C.D.S., M.F., M.A., D.M.M., G.B., S.Z., D.A.M. and A.A. collected the specimens in the field. N.I., S.M., C.D.S., M.F., J.W., G.V.L. and S.E.P. designed the research. N.I., S.M., C.D.S., M.F., J.W., G.V.L. and S.E.P. designed and performed the experiments. N.I., S.M., C.D.S., M.F., M.A., D.M.M., J.W., G.B., S.Z., D.A.M., D.M.U., U.J., J.J., A.A., G.V.L. and S.E.P. analysed the data. G.B., M.A., C.D.S., S.E.P., D.M.M., S.M., M.F. and D.B. created the figures. D.B. sculpted the life reconstruction. N.I., S.M., C.D.S., M.F., J.W., D.M.U., G.V.L. and S.E.P. wrote the manuscript, which was reviewed by all authors.

**Competing interests** The authors declare no competing interests.

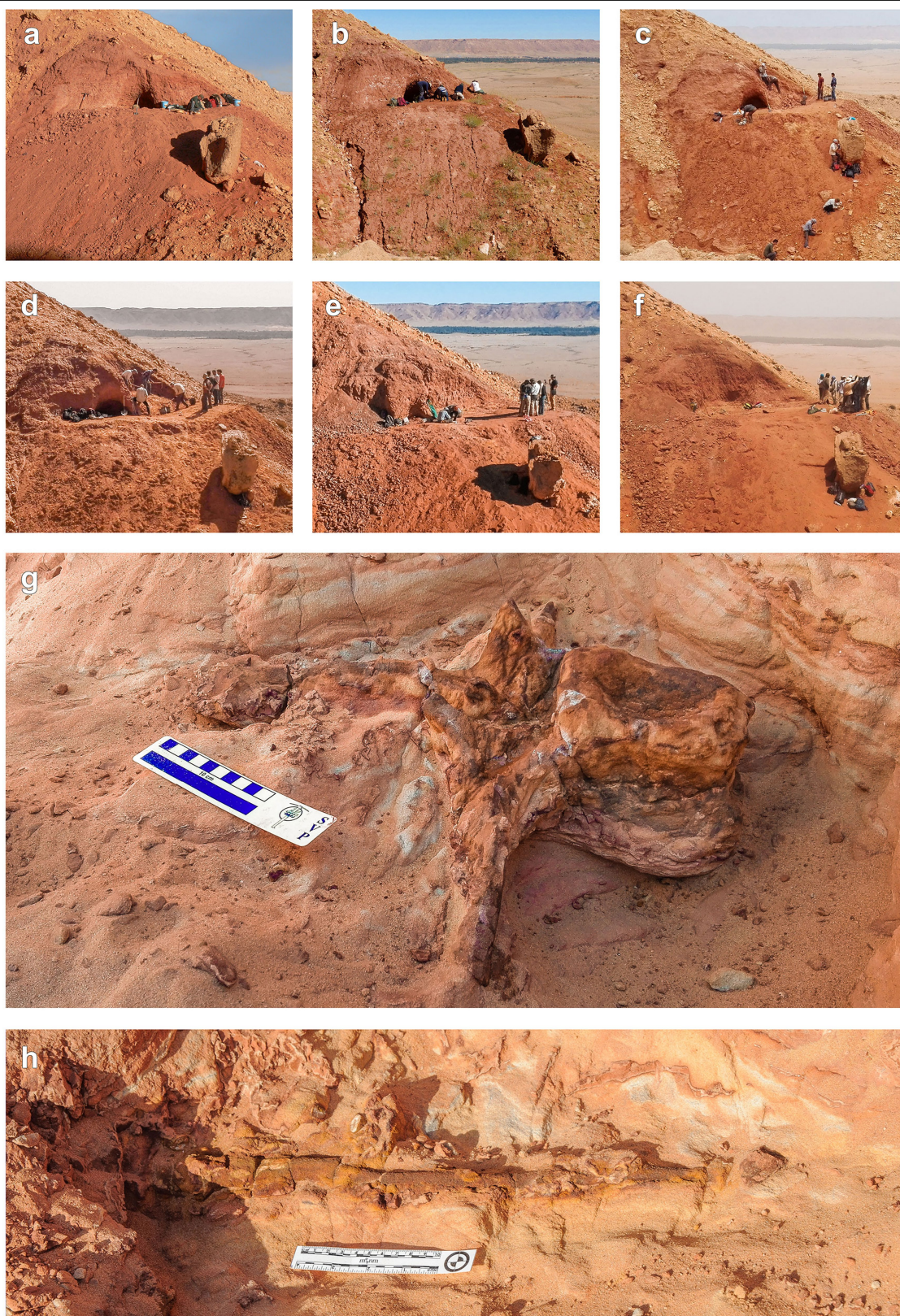
## Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41586-020-2190-3>.

**Correspondence and requests for materials** should be addressed to N.I. or S.E.P.

**Peer review information** *Nature* thanks Matthew C. Lamanna, John A. Nyakatura and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

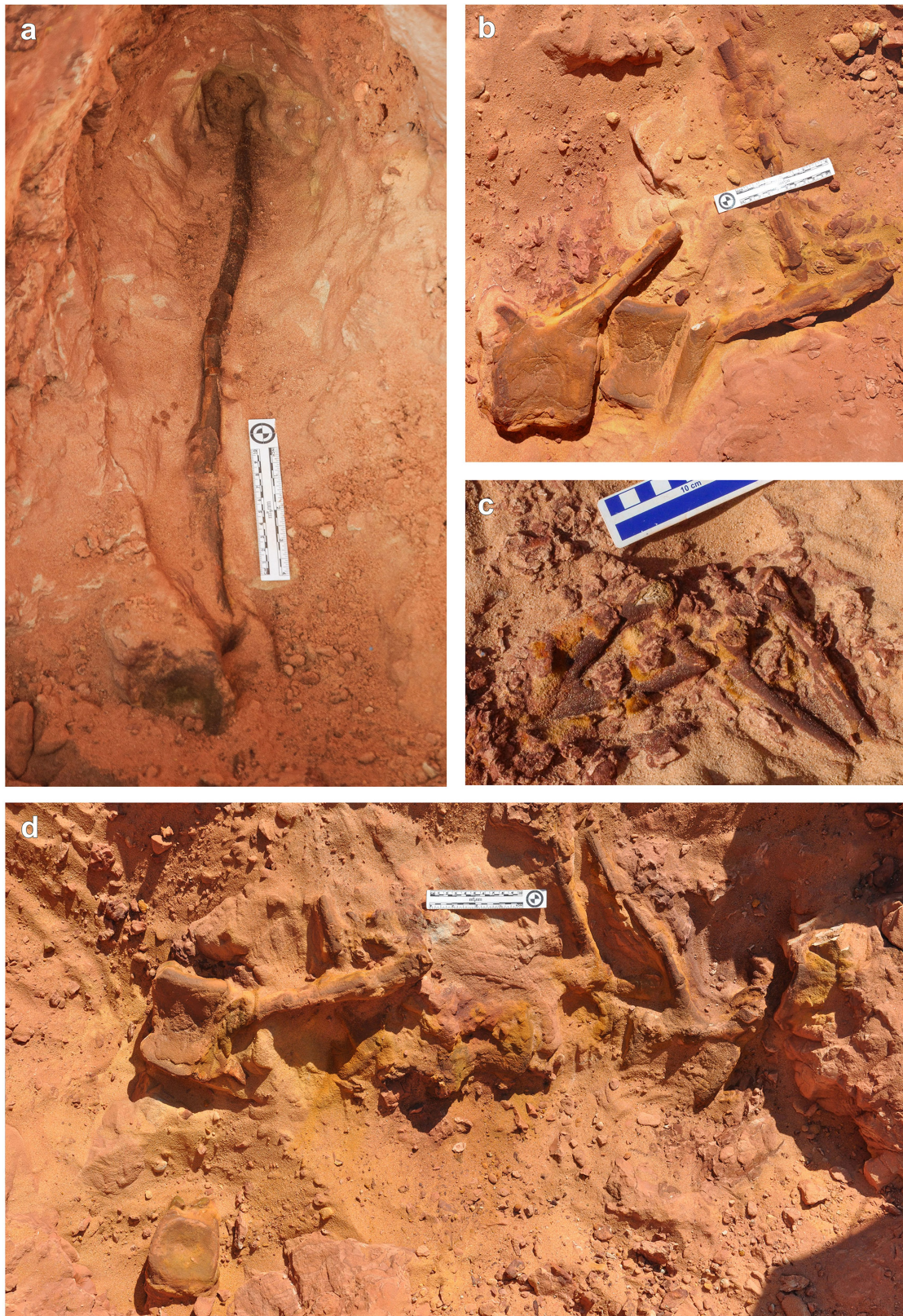
**Reprints and permissions information** is available at <http://www.nature.com/reprints>.



**Extended Data Fig. 1 | Excavation of the FSAC-KK11888 site.** **a–f**, Different stages of the excavation, which resulted in the removal of over 15 tons of rock using a range of tools, including picks, brushes, hammers and a jackhammer. **a**, 17 November 2013. **b**, 29 March 2015. **c**, 17 September 2018. **d**, 19 September

2018. **e**, 5 December 2018. **f**, 21 July 2019. **g, h**, Selected bones in situ. **g**, Largely complete proximal caudal vertebra (Ca4). **h**, Neural spine of a mid-distal caudal vertebra, fragmented by syndiagenetic cracks. Scale bars, 10 cm.

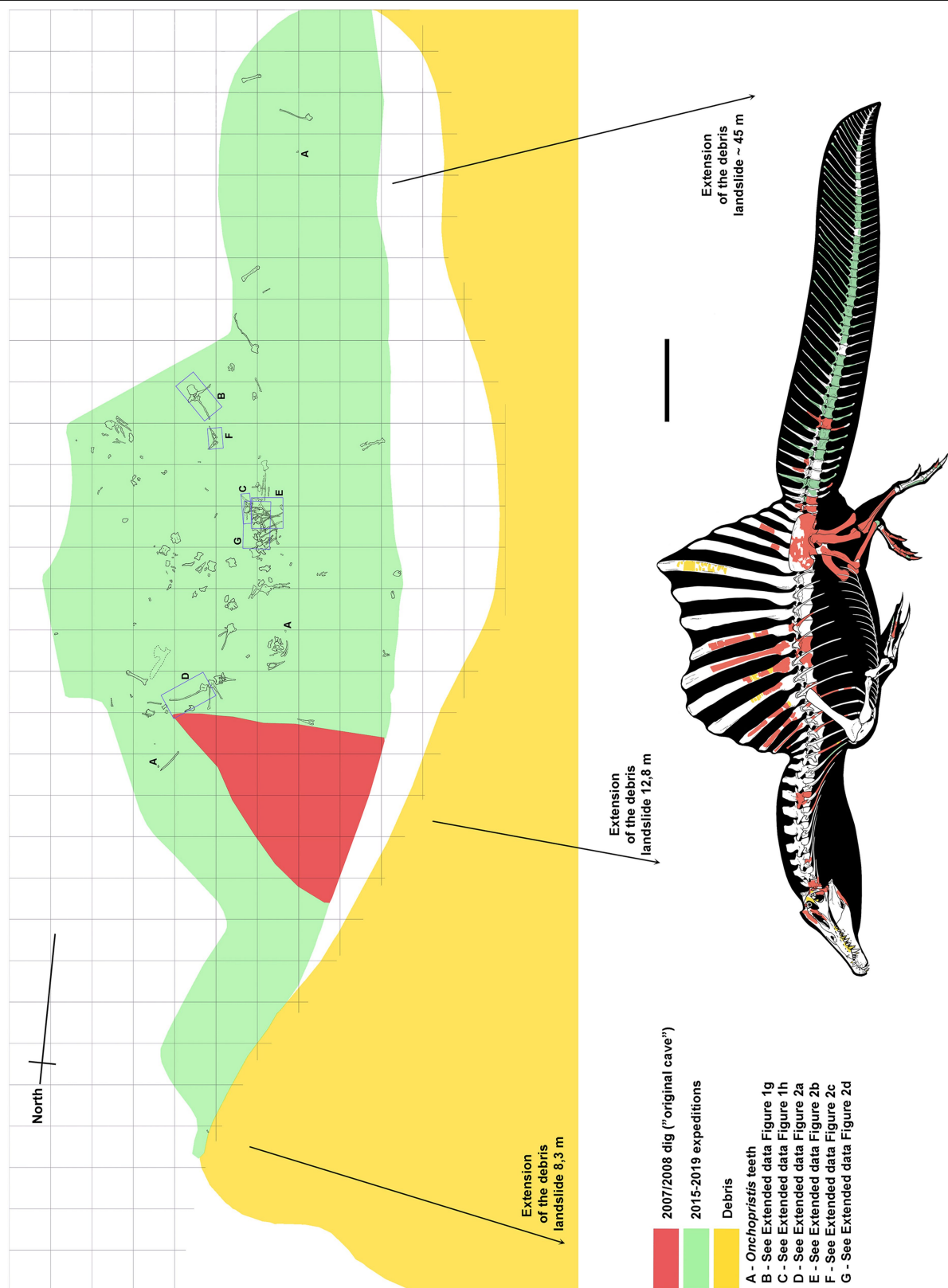




**Extended Data Fig. 2 | Excavation of the caudal elements of FSAC-KK 11888.**  
**a**, Largely complete distal caudal vertebra (Ca31), recovered in its entirety by digging a tunnel until the apex of the neural spine was reached. **b**, Semi-articulated

mid-caudal vertebrae. **c**, Two haemal arches. **d**, Close association of middle caudal elements. Scale bars, 10 cm.

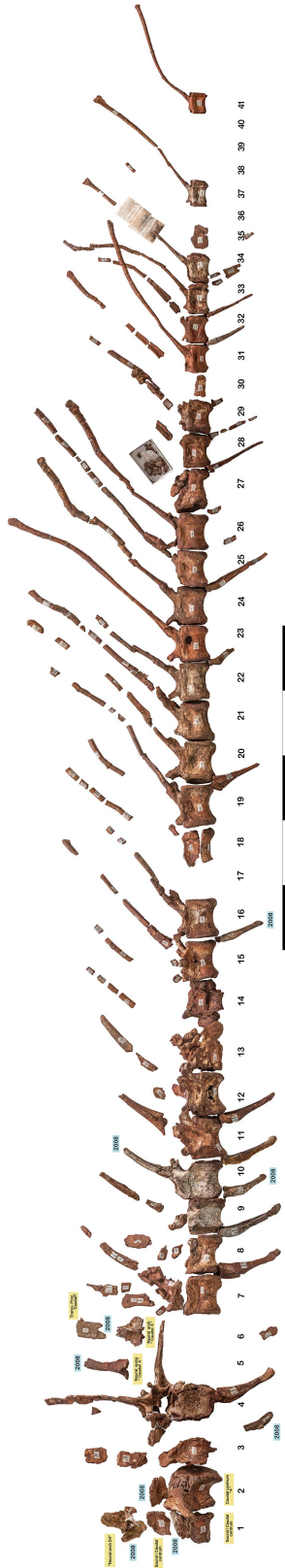




**Extended Data Fig. 3 | Excavation map and skeletal reconstruction.**  
Detailed map of the site of discovery of FSAC-KK 11888, and fully revised skeletal reconstruction. Colours in the map and reconstruction correspond to different phases of excavation: the local discovery in 2007–2008 (red), our

excavations during the 2015–2019 expeditions (green) and sieving in the debris area during the 2015–2019 expeditions (yellow). Both images are at the same scale. Scale bar, 1 m.





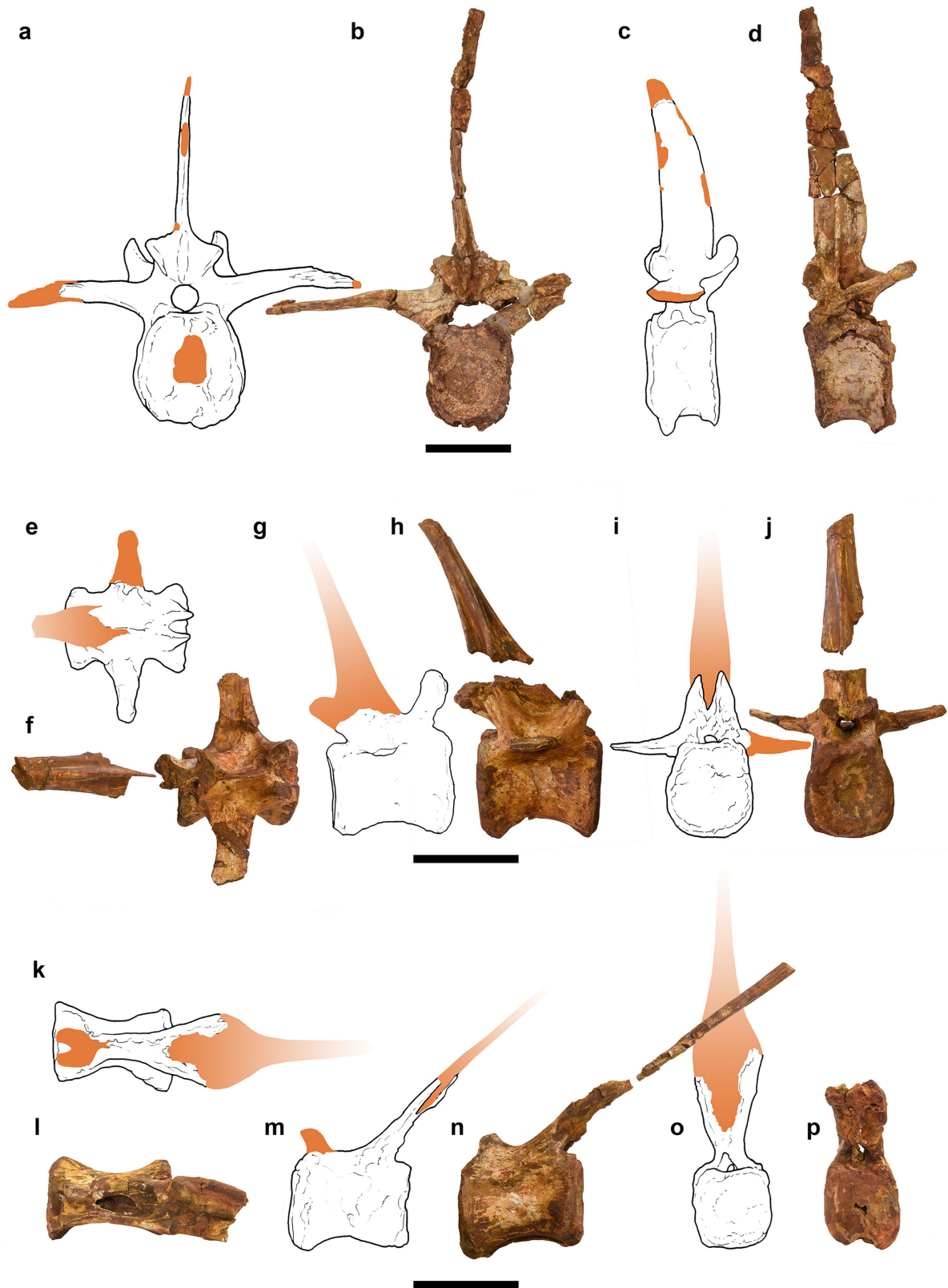
**Extended Data Fig. 4 | The caudal series of FSAC-KK 11888.** Photograph of the entire caudal series (numbered). Scale bar, 1 m.



**Extended Data Fig. 5 | Elements of FSAC-KK11888 from 2008 (first excavation) and 2019 (most recent excavation), matched. a–t,** Evidence of the perfect match between elements collected by the local discoverer of the site in 2007–2008 (a, e, f, i, k, m, n, p, q, s) and elements excavated in situ or recovered from the site debris during the 2015–2019 excavations (b–d, g, h, j, l, o, r, t). a, b, Right and left metatarsal II. c, Left penultimate phalanx of the fourth pedal digit (IV-4) that came to light within the typical matrix in which bones of the FSAC-KK11888 were embedded. d, e, Two possible splenial fragments

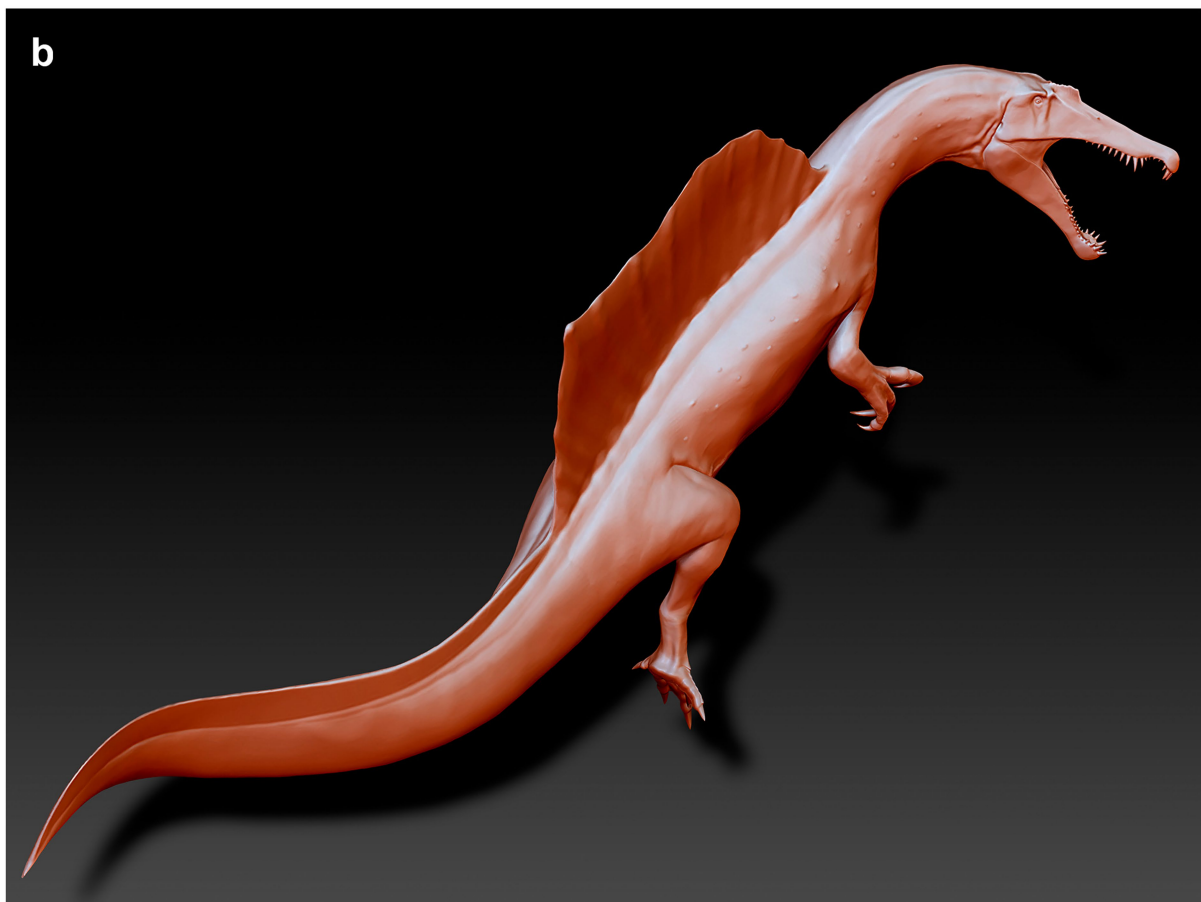
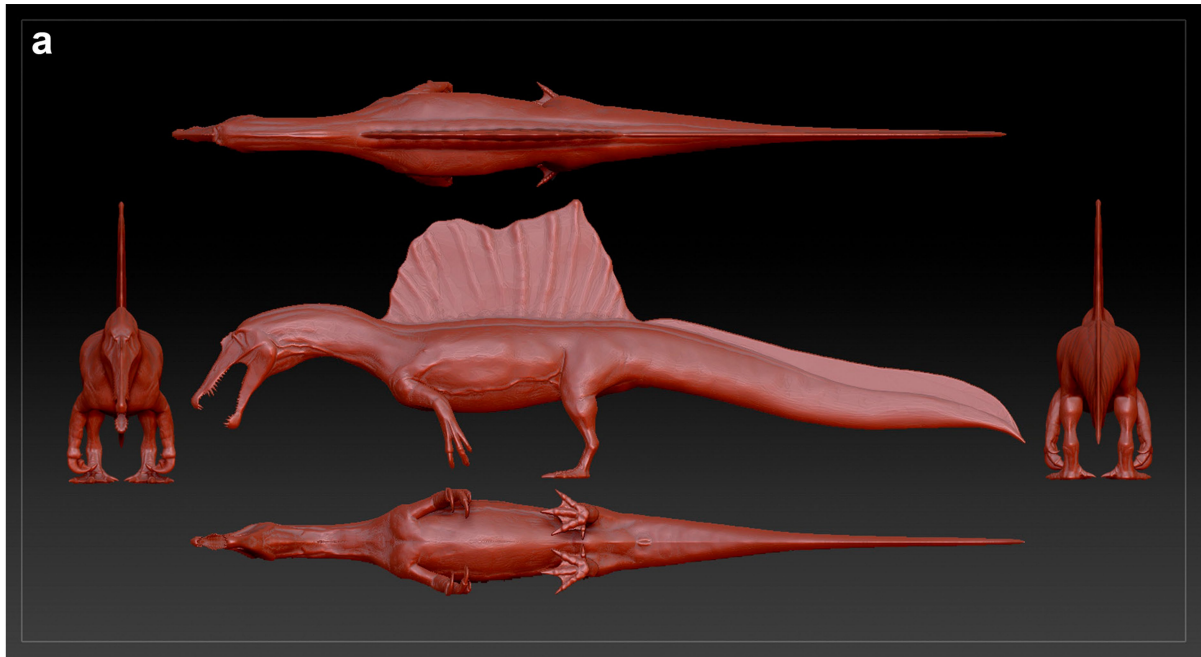
reconnected. f–i, Phalanx IV-4 prepared and compared to the contralateral element of the right pes in dorsal view, and rearticulated with its ungual. j–m, Two complementary (broken) halves of the left squamosal and of a dorsal rib. n–p, s, t, Two key fragments from the debris, reconnecting the base and the shaft of a neural spine (possibly the 7th). q, r, The right astragalus (excavated in situ in July 2019) rearticulated to its tibia (from 2008). Arrows point to recomposed fractures.





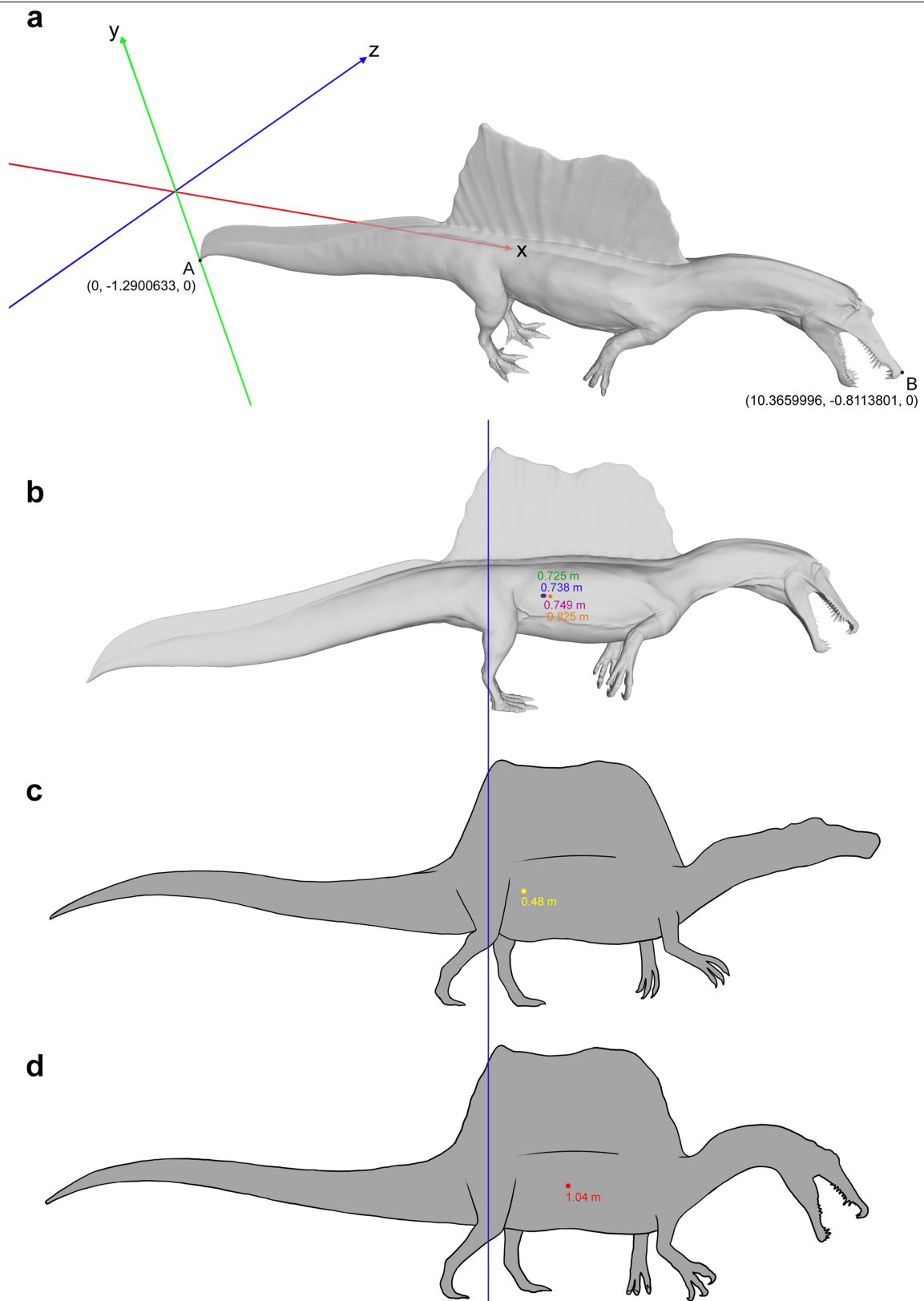
**Extended Data Fig. 6 | Comparison of FSAC-KK 11888 caudal vertebrae to those destroyed in World War II.** Comparison between the caudal vertebrae of the neotype of *S. aegyptiacus*, with those of the two specimens (the holotype and a specimen known as ‘*Spinosaurus B*’, both of which are now lost) from the Bahariya Oasis (Egypt) described by E. Stromer<sup>12</sup>. **a–d**, Proximal caudal vertebra of the holotype (accession code BSP 1912 VIII 19) in distal (**a**) and right

lateral (**c**) views; and of Ca4 of FSAC-KK 11888 in distal (**b**) and right lateral (**d**) views. **e–j**, Proximal caudal vertebra of *Spinosaurus B* in dorsal (**e**), right lateral (**g**) and proximal (**i**) views; Ca11 of FSAC-KK 11888 in dorsal (**f**), right lateral (**h**) and proximal (**j**) views. **k–p**, Middle caudal vertebra of *Spinosaurus B* in dorsal (**k**), left lateral (**m**) and distal (**o**) views; Ca21 of FSAC-KK 11888 in dorsal (**l**), left lateral (**n**) and distal (**p**) views. Scale bars = 10 cm.



**Extended Data Fig. 7 | Three-dimensional fleshed-out model based on FSAC-KK 11888. a, b, Symmetrical pose in five views (a) and swimming pose (b).**





**Extended Data Fig. 8 | Whole-body centre of mass. a–d**, Snout tip and tail tip in the coordinate system (a) and centre-of-mass distance from the cranial margin of the acetabulum in this study (b), calculated using multiple

approaches (Supplementary Information, Supplementary Data 2) and compared to the centre of mass in ref. <sup>10</sup> (c); and ref. <sup>7</sup> (d).

**Extended Data Table 1 | Measurements of caudal vertebrae of FSAC-KK 11888**

number in the series	proximodistal length of centrum	dorsoventral height proximal face of centrum	transverse diameter proximal face of centrum	minimum transverse diameter of centrum (constriction)	maximum length of neural spine (dorsal to the neural canal, along the curvature if any)	maximum width at transverse processes
Ca1	78	n.p.	n.p.	n.p.	92 (p)	320 (p, based on Ltp)
Ca2	94	129	113	67	n.p.	n.p.
Ca3	75 (p)	n.p.	n.p.	n.p.	n.p.	n.p.
Ca4	98	116	106	73	355 (p)	480 (p, based on Ltp)
Ca5	n.p.	n.p.	n.p.	n.p.	140 (p)	n.p.
Ca6	n.p.	n.p.	n.p.	n.p.	59 (p)	n.p.
Ca7	114	102	113	74	335 (p)	n.p.
Ca8	112	97	108	64	n.p.	n.p.
Ca9	99	99 (e)	81 (p)	63	n.p.	n.p.
Ca10	115 (p, L) 106 (p, R)	97 (e)	76 (p, deformed)	47	240 (p)	n.p.
Ca11	113	91	86	49	180 (p)	200 (p, almost complete)
Ca12	122	101	75	41	250 (p)	210 (based on R)
Ca13	106 (e)	79 (e)	59 (p)	33 (p)	n.p.	n.p.
Ca14	117	64 (p)	n.p.	n.p.	n.p.	n.p.
Ca15	120	84	80	47	375 (p)	186 (p, based on L)
Ca16	116	87	78	44	285 (p)	106 (p, based on L)
Ca17	n.p.	n.p.	n.p.	n.p.	210 (p)	n.p.
Ca18	102 (p)	n.p.	n.p.	n.p.	145 (p)	n.p.
Ca19	115	84	70	35	445 (p)	n.p.
Ca20	104	107	71	32	405 (p)	n.a.
Ca21	109	78	77	44	525 (p)	n.a.
Ca22	110	80	74	46	285 (p)	n.a.
Ca23	103	78	72	45	540	n.a.
Ca24	110	78	67	35	535 (p)	n.a.
Ca25	101	74	81	51	550 (p)	n.a.
Ca26	104	77	76	51	480 (p)	n.a.
Ca27	106	92	66	42	n.p.	n.a.
Ca28	102	71	65	32 (p, compressed)	n.p.	n.a.
Ca29	90	70	71	36 (p, slightly collapsed)	290 (p)	n.a.
Ca30	68 (p)	n.p.	n.p.	n.p.	295 (p)	n.a.
Ca31	86	64	58	35	535	n.a.
Ca32	85	60	58	29	325 (p)	n.a.
Ca33	91	67	51	29	505 (p)	n.a.
Ca34	87	60	62	31	290 (p)	n.a.
Ca35	75 (p)	n.p.	n.p.	n.p.	n.p.	n.a.
Ca36	n.p.	n.p.	n.p.	n.p.	n.p.	n.a.
Ca37	90	46	43	29	455	n.a.
Ca41	68	39	37	21	384	n.a.

Measurements are in mm. (p), not complete, measured as preserved; n.p., not preserved; n.a., not applicable; e, estimated.

Extended Data Table 2 | Measurements of chevrons of FSAC-KK 11888

number in the series	maximum height	height (measured perpendicular to the proximodistal axis of the tail)	base transverse width	height haemal canal
Chv3	52	n.p.	n.p.	n.p.
Chv5	91	n.p.	n.p.	n.p.
Chv6	195 (p)	170 (p)	72	74
Chv7	215	185	80	72
Chv8	204	160	73	73
Chv9	150 (p)	130 (p)	n.p.	n.p.
Chv11	195	170	81	73
Chv12	155 (p)	143 (p)	n.p.	n.p.
Chv19	160 (p)	125 (p)	72	64
Chv22	103 (p)	90 (p)	n.p.	n.p.
Chv24	224	205	65	65
Chv25	44 (p)	n.p.	n.p.	n.p.
Chv27	210	185	63	55
Chv28	155 (p)	n.p.	n.p.	n.p.
Chv29	n.p.	n.p.	n.p.	n.p.
Chv31	144 (p)	135 (p)	62	60
Chv32	176 (p) 180 (e)	160	53	53
Chv33	140 (p)	n.p.	n.p.	n.p.
Chv34	33 (p)	n.p.	n.p.	n.p.

Measurements are in mm. (p), not complete, measured as preserved; n.p., not preserved; e, estimated.

## Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

### Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

- |                                     |  |
|-------------------------------------|--|
| n/a                                 | Confirmed  |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement  |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly  |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> The statistical test(s) used AND whether they are one- or two-sided<br><i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i>  |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> A description of all covariates tested  |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons   |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For null hypothesis testing, the test statistic (e.g. $F$ , $t$ , $r$ ) with confidence intervals, effect sizes, degrees of freedom and $P$ value noted<br><i>Give <math>P</math> values as exact values whenever suitable.</i>                                       |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings  |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes  |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Estimates of effect sizes (e.g. Cohen's $d$ , Pearson's $r$ ), indicating how they were calculated  |

*Our web collection on [statistics for biologists](#) contains articles on many of the points above.*

### Software and code

Policy information about [availability of computer code](#)

Data collection	Custom LabVIEW programs (National Instruments Corp., Austin, TX, USA) for experimental testing of swimming performance.
Data analysis	Blender 2.79, ZBrush 4r7, and ClayTools 3D 2.0 for model rendering and sculpting; FlashPrint 3.26.0 to cut 3D model; Meshlab 2016.12 and Python 3.7 (through platform Anaconda) for calculation of the Centre of Mass; Meshroom 2019.1.0 for photogrammetry; custom LabVIEW 2013 programs (National Instruments Corp., Austin, TX, USA) for experimental testing of swimming performance; ImageJ 1.51w for retrocalculation of missing Lines of arrested Growth (histology).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

### Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The authors declare that all data supporting the findings of this study are available within the paper and its Supplementary Information files (Source data).



## Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☐ Behavioural & social sciences ☒ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://nature.com/documents/nr-reporting-summary-flat.pdf)

## Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	We present the first unambiguous evidence for an aquatic propulsive structure in a dinosaur, the giant theropod <i>Spinosaurus aegyptiacus</i> . This dinosaur has a tail with an unexpected and unique shape consisting of extremely tall neural spines, and elongate chevrons forming a large, flexible, fin-like organ capable of extensive lateral excursion. Using a mechanical flapping apparatus to measure undulatory forces in physical tail models, we show that the tail shape of <i>Spinosaurus</i> produces far greater thrust and efficiency than the tail shapes of terrestrial dinosaurs, comparable to that of extant aquatic vertebrates that use vertically expanded tails to generate forward propulsion while swimming. This conclusion is consistent with a suite of adaptations for an aquatic lifestyle and a piscivorous diet in <i>Spinosaurus</i> , and with a persistent and significant invasion of aquatic environments by spinosaurid dinosaurs.
Research sample	A nearly complete, partially articulated fossil tail of a subadult individual of <i>Spinosaurus aegyptiacus</i> (FSAC-KK 11888), from the Cretaceous Kem Kem beds of south-eastern Morocco. Plastic tail shapes from the following species: <i>Spinosaurus aegyptiacus</i> , <i>Coelophysis bauri</i> , <i>Allosaurus fragilis</i> , <i>Crocodylus niloticus</i> , <i>Triturus dobrogicus</i> , and a rectangular control tail scaled to the same surface area as the <i>Spinosaurus</i> tail.
Sampling strategy	Systematic collection of a partial dinosaur skeleton (one individual). Tail shape experiments: <i>Coelophysis</i> and <i>Allosaurus</i> are well known taxa and reflect a basal theropod and large-bodied theropod respectively. The two extant aquatic tetrapods (salamander and crocodile) were chosen as they are known for utilizing tail-propelled swimming.
Data collection	N.I. led the expeditions and the project. N.I., S.M., C.D.S., M.F., M.A., D.M.M., G.B., S.Z. D.M. and A.A. collected the specimens in the field. N.I., S.M., C.D.S., M.F., J.W., G.V.L. and S.E.P. designed the research. N.I., S.M., C.D.S., M.F., J.W., G.V.L. and S.E.P. designed and performed the experiments. N.I., S.M., C.D.S., M.F., M.A., D.M.M., J.W., G.B., S.Z. D.M., D.M.U., U.J., J.J., A.A., G.V.L., and S.E.P. analysed the data.
Timing and spatial scale	2018-2019; the specimen was collected from the Kem Kem beds, which crop out along the Moroccan-Algerian border for over 250 km. The fossil was collected from a well-defined field site (site map included in our submission) located near the town of Zrigat.
Data exclusions	No data were excluded
Reproducibility	The experiments were repeated over multiple days to ensure consistent measures. All repeats were successful.
Randomization	We did not use quantitative approaches that would require randomization.
Blinding	Our study (field: palaeontology) does not include experiments that would require blinding.
Did the study involve field work?	<input checked="" type="checkbox"/> Yes <input type="checkbox"/> No

## Field work, collection and transport

Field conditions	Desert escarpment close to Moroccan Algerian border
Location	In- and ex-situ on the slopes of a south-east facing escarpment fringing the Aferdou Zrigat plateau (Tafilalt basin, Akrabou Formation, Kem Kem beds). See Supplemental Information.
Access and import/export	Permits for fieldwork were obtained from Ministère de l'Energie, des Mines, et de l'Environnement. Permits: 4581/DE/2019 (issued on 17/07/2019) and 4118/DE/2018/DG (issued on 06.06.2018). The work was performed in close collaboration with researchers in Morocco (FSAC, Casablanca). The specimens collected are deposited at the Departement de Géologie/Laboratoire de Biodiversité et Santé, Faculté des Sciences Ain Chock, Hassan II University, Casablanca, Morocco.
Disturbance	No disturbance

## Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

## Materials &amp; experimental systems

n/a	Involvement in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input type="checkbox"/>	<input checked="" type="checkbox"/> Palaeontology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data

## Methods

n/a	Involvement in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

## Palaeontology

Specimen provenance	Permits for fieldwork were obtained from Ministère de l'Energie, des Mines, et de l'Environnement. Permits: 4581/DE/2019 (issued on 17/07/2019) and 4118/DE/2018/DG (issued on 06.06.2018). The work was performed in close collaboration with researchers in Morocco (FSAC, Casablanca). The specimens collected are deposited at the Departement de Géologie/Laboratoire de Biodiversité et Santé, Faculté des Sciences Ain Chock, Hassan II University, Casablanca, Morocco.
Specimen deposition	Laboratoire de Biodiversité et Santé, Faculté des Sciences Ain Chock, Hassan II University, Casablanca, Morocco
Dating methods	No new dates are provided
<input checked="" type="checkbox"/> Tick this box to confirm that the raw and calibrated dates are available in the paper or in Supplementary Information.	

# *APOE4* leads to blood–brain barrier dysfunction predicting cognitive decline

<https://doi.org/10.1038/s41586-020-2247-3>

Received: 26 November 2019

Accepted: 1 April 2020

Published online: 29 April 2020

 Check for updates

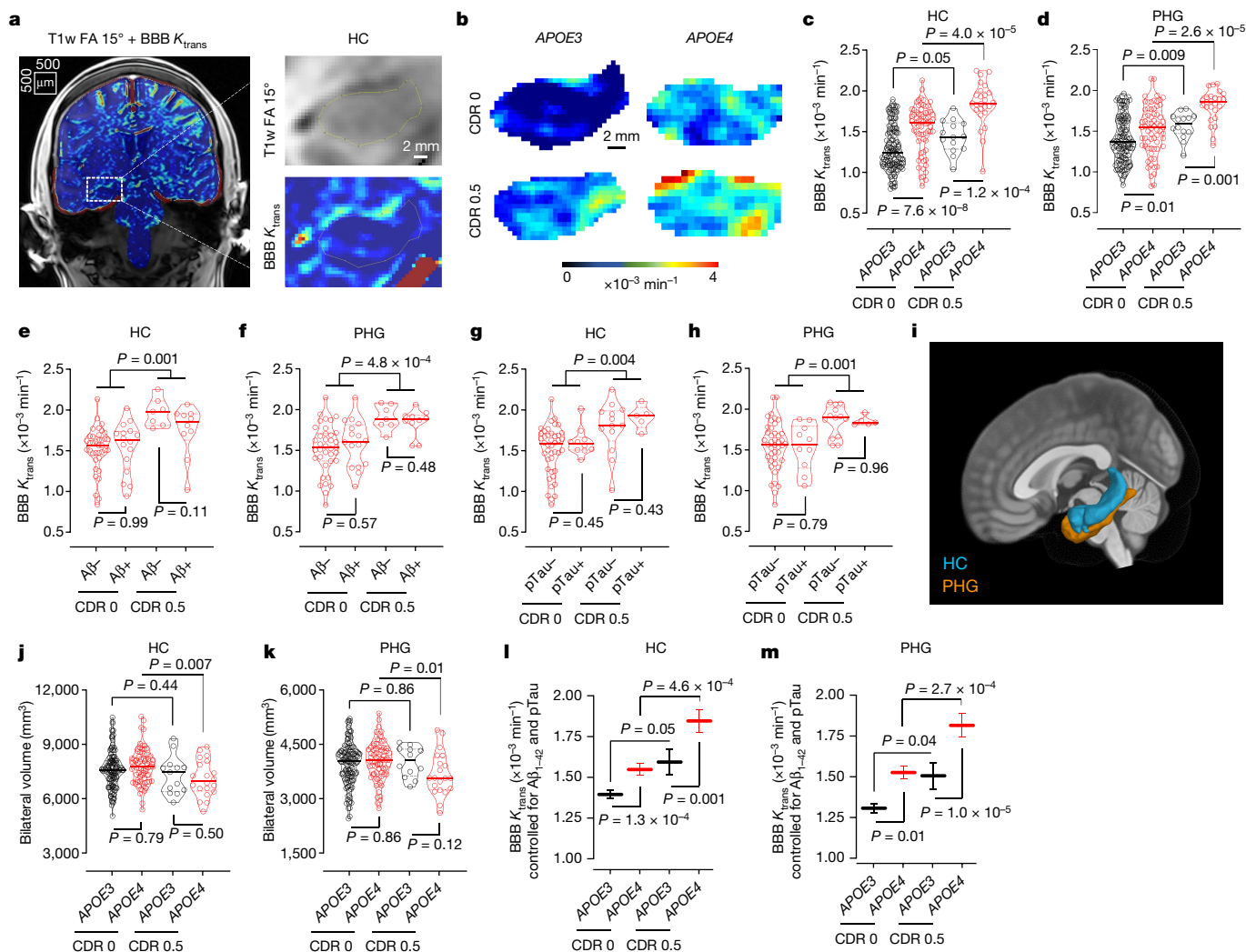
Axel Montagne<sup>1,20</sup>, Daniel A. Nation<sup>1,2,3,4,20</sup>, Abhay P. Sagare<sup>1,20</sup>, Giuseppe Barisano<sup>1,20</sup>, Melanie D. Sweeney<sup>1,20</sup>, Ararat Chakhoyan<sup>1</sup>, Maricarmen Pachicano<sup>1</sup>, Elizabeth Joe<sup>2,5</sup>, Amy R. Nelson<sup>1</sup>, Lina M. D'Orazio<sup>2,5</sup>, David P. Buennagel<sup>6</sup>, Michael G. Harrington<sup>6</sup>, Tammie L. S. Benzinger<sup>7,8</sup>, Anne M. Fagan<sup>8,9,10</sup>, John M. Ringman<sup>2,5</sup>, Lon S. Schneider<sup>2,5,11</sup>, John C. Morris<sup>9,10</sup>, Eric M. Reiman<sup>12</sup>, Richard J. Caselli<sup>13</sup>, Helena C. Chui<sup>2,5</sup>, Julia TCW<sup>14,15</sup>, Yining Chen<sup>1</sup>, Judy Pa<sup>2,16</sup>, Peter S. Conti<sup>17</sup>, Meng Law<sup>2,18,19</sup>, Arthur W. Toga<sup>2,16</sup> & Berislav V. Zlokovic<sup>1,2</sup>✉

Vascular contributions to dementia and Alzheimer's disease are increasingly recognized<sup>1–6</sup>. Recent studies have suggested that breakdown of the blood–brain barrier (BBB) is an early biomarker of human cognitive dysfunction<sup>7</sup>, including the early clinical stages of Alzheimer's disease<sup>5,8–10</sup>. The E4 variant of *apolipoprotein E* (*APOE4*), the main susceptibility gene for Alzheimer's disease<sup>11–14</sup>, leads to accelerated breakdown of the BBB and degeneration of brain capillary pericytes<sup>15–19</sup>, which maintain BBB integrity<sup>20–22</sup>. It is unclear, however, whether the cerebrovascular effects of *APOE4* contribute to cognitive impairment. Here we show that individuals bearing *APOE4* (with the  $\epsilon 3/\epsilon 4$  or  $\epsilon 4/\epsilon 4$  alleles) are distinguished from those without *APOE4* ( $\epsilon 3/\epsilon 3$ ) by breakdown of the BBB in the hippocampus and medial temporal lobe. This finding is apparent in cognitively unimpaired *APOE4* carriers and more severe in those with cognitive impairment, but is not related to amyloid- $\beta$  or tau pathology measured in cerebrospinal fluid or by positron emission tomography<sup>23</sup>. High baseline levels of the BBB pericyte injury biomarker soluble PDGFR $\beta$ <sup>7,8</sup> in the cerebrospinal fluid predicted future cognitive decline in *APOE4* carriers but not in non-carriers, even after controlling for amyloid- $\beta$  and tau status, and were correlated with increased activity of the BBB-degrading cyclophilin A-matrix metalloproteinase-9 pathway<sup>19</sup> in cerebrospinal fluid. Our findings suggest that breakdown of the BBB contributes to *APOE4*-associated cognitive decline independently of Alzheimer's disease pathology, and might be a therapeutic target in *APOE4* carriers.

Analysis of BBB permeability by dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI)<sup>24</sup> (Fig. 1a; see Methods) in 245 participants (Extended Data Table 1) indicated that there was increased BBB breakdown in the hippocampus (HC) and parahippocampal gyrus (PHG) in cognitively normal *APOE4* ( $\epsilon 3/\epsilon 4$  and  $\epsilon 4/\epsilon 4$ ) carriers, compared to cognitively normal *APOE3* homozygotes ( $\epsilon 3/\epsilon 3$ ), both with clinical dementia rating (CDR) scores of 0. The BBB breakdown in the HC and PHG in *APOE4* carriers increased further with cognitive impairment at a CDR score of 0.5 (Fig. 1b–d). This increase was independent of differences in amyloid- $\beta$  (A $\beta$ ) and phosphorylated tau (pTau) in the

cerebrospinal fluid (CSF) (Fig. 1e–h); that is, whether individuals were A $\beta$ + or A $\beta$ – and pTau+ or pTau– using the accepted cut-off values<sup>7,24,25</sup> (see Methods), where A $\beta$ + and pTau+ status indicates classical Alzheimer's disease (AD)-associated pathways<sup>23</sup>. By contrast, *APOE3* carriers with cognitive impairment developed less pronounced BBB changes in the HC and PHG (Fig. 1b–d). We found no significant BBB differences in other grey or white matter brain regions between *APOE4* carriers and *APOE3* homozygotes, except for increased BBB permeability in the caudate nucleus and minor leaks in the frontal cortex and corpus callosum in cognitively normal *APOE4* carriers (Extended Data Fig. 1).

<sup>1</sup>Department of Physiology and Neuroscience, Zilkha Neurogenetic Institute, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA. <sup>2</sup>Alzheimer's Disease Research Center, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA. <sup>3</sup>Department of Psychological Science, University of California, Irvine, Irvine, CA, USA. <sup>4</sup>Institute for Memory Disorders and Neurological Impairments, University of California, Irvine, Irvine, CA, USA. <sup>5</sup>Department of Neurology, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA. <sup>6</sup>Huntington Medical Research Institutes, Pasadena, CA, USA. <sup>7</sup>Department of Radiology, Washington University School of Medicine, St Louis, MO, USA. <sup>8</sup>The Hope Center for Neurodegenerative Disorders, Washington University School of Medicine, St Louis, MO, USA. <sup>9</sup>Department of Neurology, Washington University School of Medicine, St Louis, MO, USA. <sup>10</sup>The Knight Alzheimer's Disease Research Center, Washington University School of Medicine, St Louis, MO, USA. <sup>11</sup>Department of Psychiatry and Behavioral Sciences, University of Southern California, Los Angeles, CA, USA. <sup>12</sup>Banner Alzheimer Institute, Phoenix, AZ, USA. <sup>13</sup>Department of Neurology, Mayo Clinic, Scottsdale, AZ, USA. <sup>14</sup>Department of Neuroscience & Friedman Brain Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA. <sup>15</sup>Ronald M. Loeb Center for Alzheimer's Disease, Icahn School of Medicine at Mount Sinai, New York, NY, USA. <sup>16</sup>Laboratory of Neuroimaging, USC Stevens Neuroimaging and Informatics Institute, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA. <sup>17</sup>Molecular Imaging Center, Department of Radiology, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA. <sup>18</sup>Department of Neurological Surgery, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA. <sup>19</sup>Department of Neuroscience and Radiology, Monash University, Alfred Health, Melbourne, Victoria, Australia. <sup>20</sup>These authors contributed equally: Axel Montagne, Daniel A. Nation, Abhay P. Sagare, Giuseppe Barisano, Melanie D. Sweeney. ✉e-mail: zlokovic@usc.edu



**Fig. 1 | BBB breakdown in the HC and PHG in *APOE4* carriers increases with cognitive impairment, independently of CSF  $A\beta$  and tau status.** **a, b,** Maps of BBB permeability transfer constant ( $K_{trans}$ ) generated by DCE-MRI (**a**) in the HC of *APOE3* homozygotes (*APOE3*) and *APOE4* carriers (*APOE4*) with CDR scores of 0 or 0.5 (**b**). FA, flip angle; T1w, T1 weighted. **c, d,** BBB  $K_{trans}$  in the HC (**c**) and PHG (**d**) in individuals with CDR 0 bearing *APOE3* (black,  $n = 128$ ) or *APOE4* (red,  $n = 68$ ) and with CDR 0.5 bearing *APOE3* (black,  $n = 14$ ) or *APOE4* (red,  $n = 25$ ). **e, f,**  $K_{trans}$  in the HC (**e**) and PHG (**f**) in *APOE4* carriers with CDR 0 who were  $A\beta_{1-42}$  negative ( $A\beta^-$ ;  $n = 37$ ) or positive ( $A\beta^+$ ;  $n = 16$ ), or with CDR 0.5 who were  $A\beta^-$  ( $n = 7$ ) or  $A\beta^+$  ( $n = 10$ ). **g, h,**  $K_{trans}$  in the HC (**g**) and PHG (**h**) in *APOE4* carriers with CDR 0 who were pTau $^-$  ( $n = 42$ ) or pTau $^+$  ( $n = 10$ ), and with CDR 0.5 who were pTau $^-$  ( $n = 13$ ) or pTau $^+$  ( $n = 5$ ). **i,** HC (blue) and PHG (orange) overlaid on a 3D

template. **j, k,** Volumes of the HC (**j**) and PHG (**k**) in individuals with CDR 0 bearing *APOE3* ( $n = 124$ ) or *APOE4* ( $n = 75$ ) and with CDR 0.5 bearing *APOE3* ( $n = 13$ ) or *APOE4* ( $n = 20$ ). **l, m,**  $K_{trans}$  (estimated marginal means  $\pm$  s.e.m. from ANCOVA models corrected for age, sex, education, CSF  $A\beta_{1-42}$  and pTau status, and HC and PHG volumes) in the HC (**l**) and PHG (**m**) in individuals with CDR 0 bearing *APOE3* (black, HC  $n = 125$ ; PHG  $n = 128$ ) or *APOE4* (red, HC and PHG  $n = 68$ ) and with CDR 0.5 bearing *APOE3* (black, HC  $n = 12$ ; PHG  $n = 14$ ) or *APOE4* (red, HC  $n = 20$ ; PHG  $n = 25$ ). **c–h, j, k,** Continuous line, median; dotted line, interquartile range (IQR). Significance by ANCOVA for main effects and post hoc comparisons controlling for age, sex, and education. All ANCOVA omnibus tests remained significant at false discovery rate (FDR) threshold of 0.05.

These findings held when cognitive dysfunction was evaluated by neuropsychological performance (see Methods) (Extended Data Figs. 2, 3).

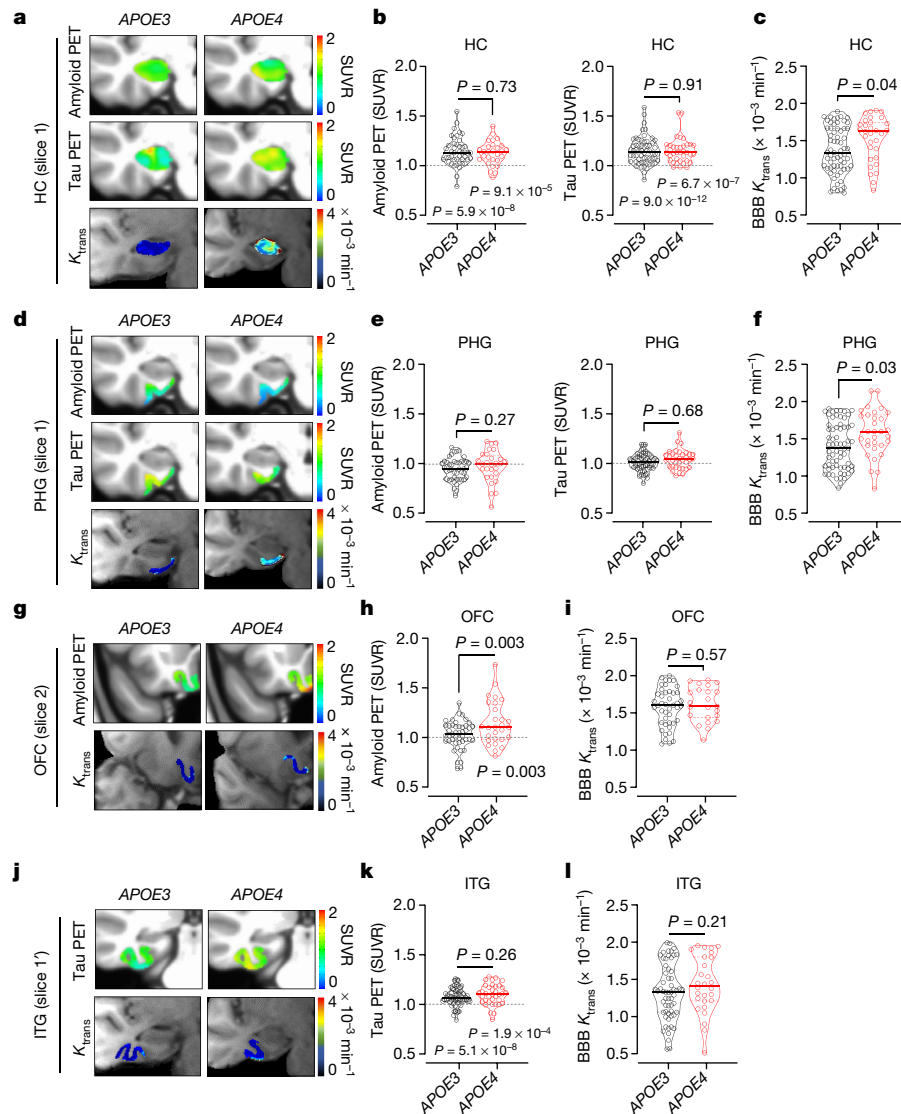
The volumes of the HC and PHG decreased with cognitive impairment in *APOE4* but not *APOE3* carriers (Fig. li–k). The breakdown of the BBB in the HC and PHG in *APOE4* carriers, but not *APOE3* homozygotes, remained a highly significant predictor of cognitive impairment after we statistically controlled for age, sex, education, CSF  $A\beta$  and pTau status, and HC and PHG volumes, as shown by the estimated marginal means from the ANCOVA models (Fig. li, m), and confirmed by logistic regression models (Supplementary Table 1). The BBB dysfunction (Fig. 1c, d, l, m) preceded brain atrophy (Fig. 1j, k) and was independent of systemic vascular risk factors (Extended Data Fig. 4).

Because both  $A\beta$  and tau can lead to blood vessel abnormalities and BBB breakdown<sup>3,26,27</sup>, we studied whether BBB disruption in *APOE4* carriers was downstream from amyloid and tau accumulation in a subset

of 74 and 96 participants, respectively (Extended Data Tables 2a, b). Voxel-based analysis by positron emission tomography (PET) indicated a substantially higher accumulation of amyloid in the orbital frontal cortex (OFC) in cognitively normal *APOE4* carriers compared to *APOE3* homozygotes, as reported<sup>28</sup>, but did not detect accumulation of tau tracer in either *APOE4* or *APOE3* carriers (Extended Data Fig. 5a–d). To determine how BBB permeability relates to accumulation of amyloid and tau, we selected 5-mm-thick coronal slices in regions of interest that included the HC and PHG (where BBB disruption is seen first in *APOE4* carriers compared to *APOE3* homozygotes (Fig. 1b, d, e)), the OFC (where amyloid accumulation develops initially in *APOE4* carriers), and the inferior temporal gyrus (ITG; a region that is affected early by tau pathology<sup>29</sup>) (Extended Data Fig. 5b, d, e).

Brain uptake of amyloid and tau tracers (after correction for the choroid plexus off-target binding for tau tracer; see Methods and





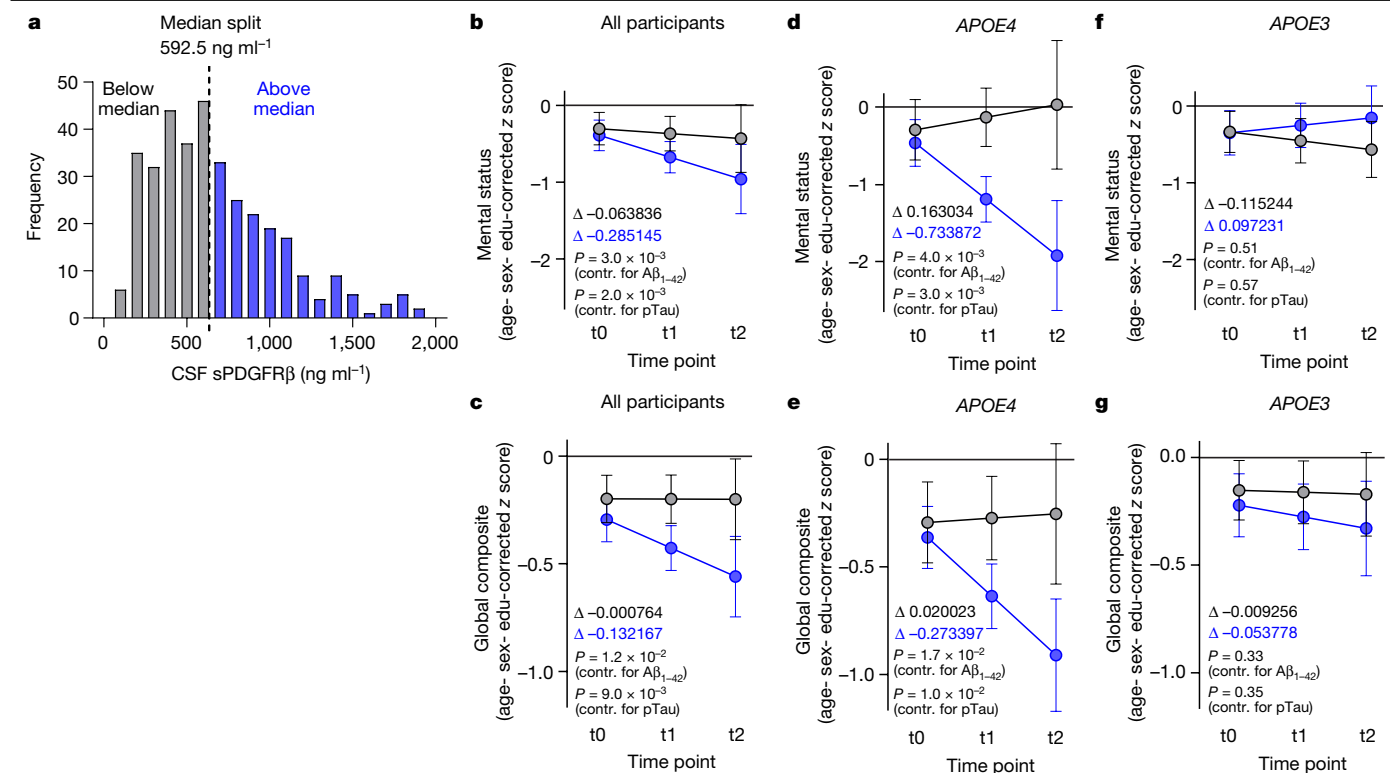
**Fig. 2 | Blood-brain barrier breakdown in *APOE4* carriers is independent of amyloid and tau accumulation in the brain.** All studies were performed in individuals with CDR score 0. **a**, Representative superimposed left HC amyloid PET (top), tau PET (middle), and BBB  $K_{trans}$  maps (bottom) from *APOE3* (left) and *APOE4* (right) carriers. **b**, **c**, Amyloid and tau tracer uptake (**b**) and BBB  $K_{trans}$  (**c**) in HC of *APOE3* ( $n = 45, 60$ , and  $65$ ) and *APOE4* ( $n = 29, 37$ , and  $31$ ) carriers. **d**, Representative superimposed left PHG amyloid PET (top), tau PET (middle), and BBB  $K_{trans}$  maps (bottom) from *APOE3* (left) and *APOE4* (right) carriers. **e**, **f**, Amyloid and tau tracer uptake (**e**) and BBB  $K_{trans}$  (**f**) in PHG of *APOE3* ( $n = 45, 60$ , and  $65$ ) and *APOE4* ( $n = 29, 37$ , and  $31$ ) carriers. **g**, Representative superimposed left medial OFC amyloid PET (top) and BBB  $K_{trans}$  maps (bottom)

from *APOE3* (left) and *APOE4* (right) carriers. **h**, **i**, Amyloid tracer uptake (**h**) and BBB  $K_{trans}$  (**i**) in OFC of *APOE3* ( $n = 45$  and  $44$ ) and *APOE4* ( $n = 29$  and  $23$ ) carriers. **j**, Representative superimposed left ITG tau PET (top) and BBB  $K_{trans}$  maps (bottom) from *APOE3* (left) and *APOE4* (right) carriers. **k**, **l**, Tau tracer uptake (**k**) and BBB  $K_{trans}$  (**l**) in ITG of *APOE3* ( $n = 60$  and  $59$ ) and *APOE4* ( $n = 37$  and  $28$ ) carriers. **b**, **c**, **e**, **f**, **h**, **i**, **k**, **l**, Continuous lines, median; dotted lines, IQR. BBB  $K_{trans}$  was determined in all participants (see Extended Data Tables 2a, b) who received either both amyloid and tau tracers ( $n = 58$ ), only amyloid tracer ( $n = 9$ ) or only tau tracer ( $n = 29$ ). Significance by ANCOVA for group comparisons controlling for age, sex, and education, and two-tailed  $t$ -tests for comparison of PET values to standardized uptake value ratios (SUVR) = 1.

Extended Data Fig. 5f, g) indicated no difference between *APOE4* and *APOE3* carriers in the HC, although uptake of both tracers was modestly increased compared to the background values in cerebellum (Fig. 2a, b). The BBB was disrupted in the HC in *APOE4* carriers compared to *APOE3* homozygotes (Fig. 2c), consistent with our findings in the larger cohort (Fig. 1b, c). There was no difference in amyloid and tau accumulation in the PHG between *APOE4* carriers and *APOE3* homozygotes, despite BBB disruption in *APOE4* carriers (Fig. 2d–f). Amyloid accumulation in the OFC was higher in cognitively normal *APOE4* carriers than in *APOE3* carriers (Fig. 2g, h), but there was no difference in BBB integrity (Fig. 2g, i). In the ITG, there were no differences in tau accumulation or BBB integrity between *APOE4* and *APOE3* carriers (Fig. 2j–l). Together,

these data suggest that BBB disruption in the HC and PHG in *APOE4* carriers is independent of AD pathology, and that BBB breakdown in *APOE4* carriers starts in the medial temporal lobe, which is responsible for memory encoding and other cognitive functions.

In humans with AD and animal models, elevated levels of soluble platelet-derived growth factor receptor- $\beta$  (sPDGFR $\beta$ ) in the CSF indicate that pericyte injury is linked to BBB breakdown<sup>7,8,30</sup> and cognitive dysfunction<sup>7,30</sup>. Using a median split for visual display of the CSF sPDGFR $\beta$  baseline levels from 350 participants (see Methods), we stratified all participants into two groups, with low CSF sPDGFR $\beta$  levels (0–600 ng ml<sup>-1</sup>) and high sPDGFR $\beta$  levels (600–2,000 ng ml<sup>-1</sup>) (Fig. 3a). In 146 *APOE4* carriers and *APOE3* homozygotes who were evaluated



**Fig. 3 | Elevated baseline CSF levels of sPDGFR $\beta$  predict cognitive decline in APOE4 carriers.** **a**, Histogram frequency distribution of CSF sPDGFR $\beta$  using median split to divide participants into two groups: high (blue, above median 600–2,000 ng ml<sup>-1</sup>) and low (grey, below median 0–600 ng ml<sup>-1</sup>) baseline CSF sPDGFR $\beta$ . All longitudinal analyses used baseline CSF sPDGFR $\beta$  as a continuous predictor of future cognitive decline. **b**, **c**, Linear mixed model analysis of all studied participants ( $n = 146$ ) followed at 2-year intervals for up to 4.5 years after baseline lumbar puncture (LP). Higher baseline CSF sPDGFR $\beta$  (blue) predicts greater decline in demographically corrected mental status exam scores over time ( $P = 0.01$ ) (this remains significant after controlling (contr.) for CSF A $\beta$  ( $P = 0.002$ ) and pTau ( $P = 0.002$ ) status; **b**), and in global cognitive composite scores ( $P = 0.01$ ) (this remains significant after controlling for CSF A $\beta$  ( $P = 0.017$ ) and pTau ( $P = 0.01$ ) status; **c**). **d**, **e**, Higher CSF sPDGFR $\beta$  (blue) in

APOE4 carriers ( $n = 58$ ) predicts future decline in mental status exam scores ( $P = 0.005$ ) after controlling for CSF A $\beta$  ( $P = 0.004$ ) and pTau ( $P = 0.003$ ) status (**d**), and in global cognitive composite scores ( $P = 0.02$ ) after controlling for CSF A $\beta$  ( $P = 0.02$ ) and pTau ( $P = 0.01$ ) status (**e**). **f**, **g**, Baseline CSF sPDGFR $\beta$  does not predict decline ( $n = 88$ ) in either mental status (**f**) or global composite (**g**) scores in APOE3 homozygotes regardless of CSF A $\beta$  or pTau status. **b**–**g**, Separate lines indicate median split of baseline CSF sPDGFR $\beta$  (grey, below median; blue, above median).  $\Delta$  slopes provided for median split of baseline CSF sPDGFR $\beta$  groups. t0 = –1 to 0.5 years post-LP, t1 = 0.5 to 2.5 years post-LP, t2 = 2.5 to 4.5 years post-LP. Error bars show s.e. of the estimate. Linear mixed model analysis with no multiple comparison. See Supplementary Tables 2–4 for detailed statistics.

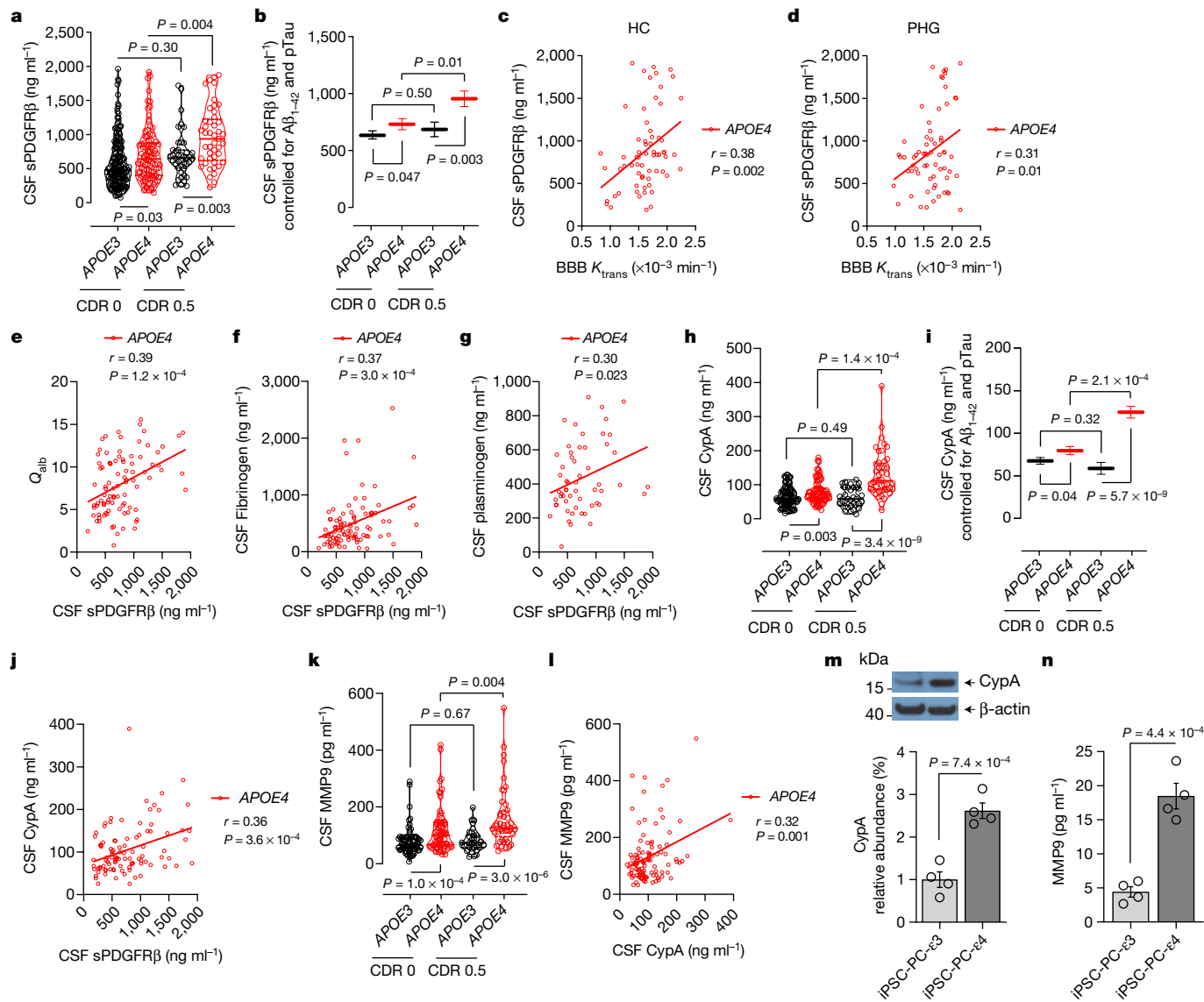
by cognitive exams at 2-year intervals up to 4.5 years from baseline, participants with higher baseline CSF sPDGFR $\beta$  exhibited accelerated cognitive decline on a global mental status exam and global cognitive composite z-scores, which remained significant after controlling for CSF A $\beta$  and tau status (Fig. 3b, c; Supplementary Table 2). When stratified by APOE status, higher baseline CSF sPDGFR $\beta$  levels in APOE4 carriers predicted cognitive decline after controlling for CSF A $\beta$  and pTau status (Fig. 3d, e; Supplementary Table 3), but did not predict decline in APOE3 homozygotes (Fig. 3f, g; Supplementary Table 4).

The increase in CSF sPDGFR $\beta$  with cognitive impairment was also found on cross-sectional CDR analysis in APOE4 carriers but not APOE3 homozygotes (Fig. 4a, b; Extended Data Table 3; Supplementary Table 5). Increased levels of sPDGFR $\beta$  in the CSF of APOE4 carriers correlated with increases in BBB permeability in the HC and PHG (Fig. 4c, d), and elevated levels of molecular biomarkers of BBB breakdown including albumin CSF/plasma quotient, and CSF fibrinogen and plasminogen (Fig. 4e–g).

Next, we focused on the proinflammatory cyclophilin A–matrix metalloproteinase-9 (CypA–MMP9) pathway. When activated by brain capillary pericytes in APOE4 (but not APOE3) knock-in mice, this pathway leads to MMP9-mediated breakdown of the BBB, which in turn induces neuronal stress related to leaked blood-derived neurotoxic proteins followed by neuronal dysfunction and loss of synaptic proteins<sup>19</sup>.

Brain tissue analysis has also shown higher activation of the CypA–MMP9 pathway in degenerating brain capillary pericytes in APOE4 carriers than in APOE3 homozygotes<sup>16</sup>. In our cohort, APOE4 carriers, but not APOE3 homozygotes, developed an increase in CypA CSF levels with cognitive impairment (Fig. 4h, i), which correlated with elevated CSF sPDGFR $\beta$  (Fig. 4j). APOE4 carriers, but not APOE3 homozygotes, also developed elevated MMP9 in the CSF with cognitive impairment (Fig. 4k), which correlated with elevated CSF CypA levels (Fig. 4l), suggesting that activation of the CypA–MMP9 pathway in APOE4 carriers correlates with pericyte injury, as shown in animal models<sup>19</sup>. There were no differences in glia or in inflammatory or endothelial cell injury CSF biomarkers between cognitively impaired and unimpaired APOE4 and APOE3 participants, but there was an increase in neuron-specific enolase (NSE) with cognitive impairment in APOE4 carriers, confirming neuronal stress (Extended Data Fig. 6) and consistent with atrophy of the HC and PHG (Fig. 1j, k).

Studies in APOE knock-in mice and mouse pericytes have shown that apoE3, but not apoE4, transcriptionally inhibits CypA via low-density lipoprotein receptor-related protein 1, which in turn transcriptionally inhibits MMP9<sup>19</sup>. Consistent with the mouse data, pericytes derived from APOE4 ( $\epsilon 4/\epsilon 4$ ) human induced pluripotent stem cells (iPSCs) had substantially higher levels of CypA and secreted MMP9 than those derived from APOE3 ( $\epsilon 3/\epsilon 3$ ) cells (Fig. 4m, n), suggesting that



**Fig. 4 | Elevated CSF sPDGFR $\beta$ , cyclophilin A and matrix metalloproteinase-9 in *APOE4* carriers.** **a**, CSF sPDGFR $\beta$  in individuals with CDR 0 bearing *APOE3* (black,  $n = 152$ ) or *APOE4* (red,  $n = 95$ ) and with CDR 0.5 bearing *APOE3* (black,  $n = 42$ ) or *APOE4* (red,  $n = 45$ ). **b**, CSF sPDGFR $\beta$  (estimated marginal means  $\pm$  s.e.m. from ANCOVA models corrected for age, sex, education, CSF  $A\beta_{1-42}$  and pTau status) in individuals with CDR 0 bearing *APOE3* (black,  $n = 152$ ) or *APOE4* (red,  $n = 95$ ) and with CDR 0.5 bearing *APOE3* (black,  $n = 42$ ) or *APOE4* (red,  $n = 45$ ). **c**, **d**, Correlation between CSF sPDGFR $\beta$  and BBB  $K_{trans}$  in the HC ( $n = 65$ ; **c**) and PHG ( $n = 65$ ; **d**) in *APOE4* carriers. **e**–**g**, Correlations between CSF sPDGFR $\beta$  and albumin quotient ( $Q_{alb}$ ,  $n = 92$ ; **e**), fibrinogen ( $n = 93$ ; **f**), and plasminogen ( $n = 57$ ; **g**) in *APOE4* carriers. **h**, CSF CypA in individuals with CDR 0 bearing *APOE3* (black,  $n = 75$ ) or *APOE4* (red,  $n = 62$ ) and with CDR 0.5 bearing *APOE3* (black,  $n = 33$ ) or *APOE4* (red,  $n = 45$ ). **i**, CSF CypA (estimated marginal means  $\pm$  s.e.m. as in **b**) in individuals with CDR 0 bearing *APOE3* (black,  $n = 75$ ) or

*APOE4* (red,  $n = 62$ ) and with CDR 0.5 bearing *APOE3* (black,  $n = 33$ ) or *APOE4* (red,  $n = 45$ ). **j**, Correlation between CSF CypA and sPDGFR $\beta$  in *APOE4* carriers ( $n = 96$ ). **k**, CSF matrix metalloproteinase-9 (MMP9) in individuals with CDR 0 bearing *APOE3* (black,  $n = 72$ ) or *APOE4* (red,  $n = 68$ ) and with CDR 0.5 bearing *APOE3* (black,  $n = 33$ ) or *APOE4* (red,  $n = 45$ ). **l**, Correlation between CSF MMP9 and CypA in *APOE4* carriers ( $n = 104$ ). **m**, **n**, CypA (**m**; see Extended Data Fig. 8) and secreted MMP9 in the culture medium (**n**) in human iPSC-derived *APOE3* ( $\epsilon3/\epsilon3$ ) and *APOE4* ( $\epsilon4/\epsilon4$ ) pericytes. Mean  $\pm$  s.e.m. from four independent culture replicates. **a**, **h**, **k**, Continuous lines, median; dotted lines, IQR. Significance by ANCOVA for main effects and post hoc comparisons controlling for age, sex, and education. **c**–**j**, **l**, Two-tailed simple linear regression; Pearson correlation coefficient ( $r$ ). **m**, **n**, Unpaired two-tailed Student's  $t$ -test.

apoE may control the CypA–MMP9 pathway in human pericytes in an isoform-specific manner, as in mouse models<sup>19</sup>.

In *APOE4* carriers, CSF  $A\beta_{1-42}$  was reduced and CSF pTau levels were increased with cognitive impairment, compared to *APOE3* homozygotes (Extended Data Fig. 7), as reported<sup>23</sup>; this difference remained significant after controlling for CSF sPDGFR $\beta$  levels (Extended Data Fig. 7). Together, these findings support the idea that the  $A\beta$  and tau pathways operate independently of the BBB breakdown pathway during the early stages of cognitive impairment in *APOE4* carriers.

In summary, we have shown that BBB breakdown contributes to cognitive decline in *APOE4* carriers independent of AD pathology; that high baseline CSF levels of sPDGFR $\beta$  can predict future cognitive decline in *APOE4* carriers; and that *APOE4*, but not *APOE3*, activates the CypA–MMP9 pathway in the CSF, which may lead to accelerated BBB breakdown and thereby cause neuronal and synaptic dysfunction<sup>19</sup>. As blockade of the CypA–MMP9 pathway in *APOE4* knock-in mice restores BBB integrity and subsequently normalizes neuronal and synaptic function<sup>19</sup>, it is possible that CypA inhibitors (some of which have been used

in humans for non-neurological applications<sup>31</sup>) might also suppress the CypA pathway in cerebral blood vessels in *APOE4* carriers. This should improve cerebrovascular integrity, and reduce the associated neuronal and synaptic deficits, thereby slowing cognitive impairment.

## Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-020-2247-3>.

- Wardlaw, J. M. et al. Neuroimaging standards for research into small vessel disease and its contribution to ageing and neurodegeneration. *Lancet Neurol.* **12**, 822–838 (2013).
- Kapasi, A., DeCarli, C. & Schneider, J. A. Impact of multiple pathologies on the threshold for clinically overt dementia. *Acta Neuropathol.* **134**, 171–186 (2017).
- Iadecola, C. The neurovascular unit coming of age: a journey through neurovascular coupling in health and disease. *Neuron* **96**, 17–42 (2017).
- Iturria-Medina, Y., Sotero, R. C., Toussaint, P. J., Mateos-Pérez, J. M. & Evans, A. C. Early role of vascular dysregulation on late-onset Alzheimer's disease based on multifactorial data-driven analysis. *Nat. Commun.* **7**, 11934 (2016).
- Sweeney, M. D., Sagare, A. P. & Zlokovic, B. V. Blood–brain barrier breakdown in Alzheimer disease and other neurodegenerative disorders. *Nat. Rev. Neurol.* **14**, 133–150 (2018).
- Sweeney, M. D. et al. Vascular dysfunction—the disregarded partner of Alzheimer's disease. *Alzheimers Dement.* **15**, 158–167 (2019).
- Nation, D. A. et al. Blood–brain barrier breakdown is an early biomarker of human cognitive dysfunction. *Nat. Med.* **25**, 270–276 (2019).
- Montagne, A. et al. Blood–brain barrier breakdown in the aging human hippocampus. *Neuron* **85**, 296–302 (2015).
- van de Haar, H. J. et al. Neurovascular unit impairment in early Alzheimer's disease measured with magnetic resonance imaging. *Neurobiol. Aging* **45**, 190–196 (2016).
- van de Haar, H. J. et al. Blood–brain barrier leakage in patients with early Alzheimer disease. *Radiology* **281**, 527–535 (2016).
- Corder, E. H. et al. Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. *Science* **261**, 921–923 (1993).
- Roses, A. D. Apolipoprotein E alleles as risk factors in Alzheimer's disease. *Annu. Rev. Med.* **47**, 387–400 (1996).
- Farrer, L. A. et al. Effects of age, sex, and ethnicity on the association between apolipoprotein E genotype and Alzheimer disease. A meta-analysis. *J. Am. Med. Assoc.* **278**, 1349–1356 (1997).
- Genin, E. et al. APOE and Alzheimer disease: a major gene with semi-dominant inheritance. *Mol. Psychiatry* **16**, 903–907 (2011).
- Hultman, K., Strickland, S. & Norris, E. H. The APOE  $\epsilon 4/\epsilon 4$  genotype potentiates vascular fibrin(ogen) deposition in amyloid-laden vessels in the brains of Alzheimer's disease patients. *J. Cereb. Blood Flow Metab.* **33**, 1251–1258 (2013).
- Halliday, M. R. et al. Accelerated pericyte degeneration and blood–brain barrier breakdown in apolipoprotein E4 carriers with Alzheimer's disease. *J. Cereb. Blood Flow Metab.* **36**, 216–227 (2016).
- Salloway, S. et al. Effect of APOE genotype on microvascular basement membrane in Alzheimer's disease. *J. Neurol. Sci.* **203–204**, 183–187 (2002).
- Zipser, B. D. et al. Microvascular injury and blood–brain barrier leakage in Alzheimer's disease. *Neurobiol. Aging* **28**, 977–986 (2007).
- Bell, R. D. et al. Apolipoprotein E controls cerebrovascular integrity via cyclophilin A. *Nature* **485**, 512–516 (2012).
- Armulik, A. et al. Pericytes regulate the blood–brain barrier. *Nature* **468**, 557–561 (2010).
- Bell, R. D. et al. Pericytes control key neurovascular functions and neuronal phenotype in the adult brain and during brain aging. *Neuron* **68**, 409–427 (2010).
- Nikolakopoulou, A. M. et al. Pericyte loss leads to circulatory failure and pleiotrophin depletion causing neuron loss. *Nat. Neurosci.* **22**, 1089–1098 (2019).
- Jack, C. R. Jr et al. NIA-AA Research Framework: toward a biological definition of Alzheimer's disease. *Alzheimers Dement.* **14**, 535–562 (2018).
- Pan, C. et al. Diagnostic values of cerebrospinal fluid T-Tau and A $\beta_{42}$  using meso scale discovery assays for Alzheimer's disease. *J. Alzheimers Dis.* **45**, 709–719 (2015).
- Roe, C. M. et al. Amyloid imaging and CSF biomarkers in predicting cognitive impairment up to 7.5 years later. *Neurology* **80**, 1784–1791 (2013).
- Montagne, A., Zhao, Z. & Zlokovic, B. V. Alzheimer's disease: a matter of blood–brain barrier dysfunction? *J. Exp. Med.* **214**, 3151–3169 (2017).
- Bennett, R. E. et al. Tau induces blood vessel abnormalities and angiogenesis-related gene expression in P301L transgenic mice and human Alzheimer's disease. *Proc. Natl Acad. Sci. USA* **115**, E1289–E1298 (2018).
- Fouquet, M., Besson, F. L., Gonneaud, J., La Joie, R. & Chételat, G. Imaging brain effects of APOE4 in cognitively normal individuals across the lifespan. *Neuropsychol. Rev.* **24**, 290–299 (2014).
- Schultz, S. A. et al. Widespread distribution of tauopathy in preclinical Alzheimer's disease. *Neurobiol. Aging* **72**, 177–185 (2018).
- Miners, J. S., Kehoe, P. G., Love, S., Zetterberg, H. & Blennow, K. CSF evidence of pericyte damage in Alzheimer's disease is associated with markers of blood–brain barrier dysfunction and disease pathology. *Alzheimers Res. Ther.* **11**, 81 (2019).
- Stanciu, C., Trifan, A., Muzica, C. & Sfarti, C. Efficacy and safety of alisporivir for the treatment of hepatitis C infection. *Expert Opin. Pharmacother.* **20**, 379–384 (2019).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2020



## Methods

### Study participants

Participants were recruited from three sites: the University of Southern California (USC), Los Angeles, CA; Washington University (WashU), St. Louis, MO; and Banner Alzheimer's Institute Phoenix, AZ and Mayo Clinic Arizona, Scottsdale, AZ as a single site. At the USC site, participants were recruited through the USC Alzheimer's Disease Research Center (ADRC): combined USC and the Huntington Medical Research Institutes (HMRI), Pasadena, CA. At the WashU site, participants were recruited through the Washington University Knight ADRC. At Banner Alzheimer's Institute and Mayo Clinic Arizona site, participants were recruited through the Arizona Apolipoprotein E (*APOE*) cohort. The study and procedures were approved by the Institutional Review Boards of USC ADRC, Washington University Knight ADRC, and Banner Good Samaritan Medical Center and Mayo Clinic Scottsdale, indicating compliance with all ethical regulations. Informed consent was obtained from all participants before study enrolment. All participants ( $n = 435$ ) underwent neurological and neuropsychological evaluations performed using the Uniform Data Set (UDS)<sup>32</sup> and additional neuropsychological tests, as described below, and received a venipuncture for collection of blood for biomarker studies. An LP was performed in 350 participants (81%) for collection of CSF. DCE-MRI for assessment of BBB permeability was performed in 245 participants (56%) who had no contraindications for contrast injection. Both LP and DCE-MRI were conducted in 172 participants. Among the 245 DCE-MRI participants, 74 and 96 were additionally studied for brain uptake of amyloid and tau PET radiotracers, respectively, as described below. No statistical methods were used to predetermine sample size. All biomarker assays, MRI, and PET scans were analysed by investigators blinded to the clinical status of the participants.

### Participant inclusion and exclusion criteria

Included participants ( $\geq 45$  years of age) were confirmed by clinical and cognitive assessments to be either cognitively normal or at the earliest symptomatic stage of AD. A current or prior history of any neurological or psychiatric conditions that might confound cognitive assessment, including organ failure, brain tumours, epilepsy, hydrocephalus, schizophrenia, and major depression, was exclusionary. Participants were stratified by *APOE* genotype as *APOE4* carriers ( $\epsilon 3/\epsilon 4$  and  $\epsilon 4/\epsilon 4$ ) or *APOE4* non-carriers ( $\epsilon 3/\epsilon 3$ ), also defined as *APOE3* homozygotes, who were cognitively normal or had mild cognitive dysfunction, as determined by CDR scores<sup>33</sup> and the presence of cognitive impairment in one or more cognitive domains based on comprehensive neuropsychological evaluation, including performance on ten neuropsychological tests assessing memory, attention/executive function, language and global cognition. For all analyses individuals with  $\epsilon 3/\epsilon 4$  and  $\epsilon 4/\epsilon 4$  alleles were pooled together in a single *APOE4* group, as we did not find in the present cohort (82–86%  $\epsilon 3/\epsilon 4$  and 14–18%  $\epsilon 4/\epsilon 4$  participants, depending on the outcome measure) a significant difference between individuals with two versus one  $\epsilon 4$  allele for the studied parameters, including the BBB  $K_{trans}$  and sPDGFR $\beta$  CSF values (see statistical section below). Individuals were additionally stratified by  $A\beta$  and pTau CSF analysis as either  $A\beta_{1-42}^+$  ( $<190$  pg/ml) or  $A\beta_{1-42}^-$  ( $>190$  pg/ml), and pTau+ ( $>78$  pg/ml) or pTau- ( $<78$  pg/ml), using accepted cutoff values<sup>7,24,25</sup>.

Participants were excluded if they were diagnosed with vascular cognitive impairment or vascular dementia. Clinical diagnoses were made by neurologists and criteria included whether the patient had a known vascular brain injury, and whether the clinician judged that the vascular brain injury played a role in their cognitive impairment, and/or pattern and course of symptoms. In addition to clinical diagnosis, the presence of vascular lesions was confirmed by moderate-to-severe white matter changes and lacunar infarcts by fluid-attenuated inversion recovery (FLAIR) MRI and/or subcortical microbleeds by T2\*-weighted MRI<sup>1</sup>. Participants were also excluded if they were diagnosed with Parkinson's

disease, Lewy body dementia or frontotemporal dementia. History of a single stroke or transient ischaemic attack was not an exclusion unless it was related to symptomatic onset of cognitive impairment. Participants also did not have current contraindications to MRI and were not currently using medications that might better account for any observed cognitive impairment.

### Clinical exam

Participants underwent clinical assessments according to UDS procedures harmonized across all study sites, including clinical interview and review of any neurocognitive symptoms and health history with the participant and a knowledgeable informant. A general physical and neurologic exam was conducted. The CDR assessment was conducted in accordance with published standardization procedures, including standardized interview and assessment with the participant and a knowledgeable informant. In accordance with current diagnostic models for cognitive and biological research criteria for cognitive impairment and AD<sup>23</sup>, participants were separately stratified by cognitive impairment and AD biomarker abnormality using established cutoffs for CSF  $A\beta_{1-42}$  and pTau<sup>7,24,25</sup>. Cognitive impairment was determined on the basis of global CDR score and neuropsychological impairment in one or more cognitive domains.

### Vascular risk factors

The vascular risk factor (VRF) burden in each participant was evaluated through physical examination, blood tests, and clinical interviews with the participant and informant; history of cardiovascular disease (heart failure, angina, stent placement, coronary artery bypass graft, intermittent claudication), hypertension, hyperlipidaemia, type 2 diabetes, atrial fibrillation, and transient ischaemic attack or minor stroke were investigated. The total VRF burden was defined by the sum of these risk factors, as previously described<sup>7</sup>. We assigned an elevated VRF burden to individuals with two or more VRFs. This threshold was adopted because previous studies showed that the presence of two or more VRFs is associated with occult cerebrovascular disease at autopsy in older adults with AD, whereas a single VRF is common and not necessarily associated with increased cerebrovascular disease in this population<sup>34,35</sup>.

### Cognitive domain impairment evaluation

Impairment in one or more cognitive domain was judged by performance on comprehensive neuropsychological testing, using previously described neuropsychological criteria for cognitive impairment<sup>7</sup>. All participants underwent neuropsychological testing that included the UDS battery (version 2.0 or 3.0) plus supplementary neuropsychological tests at each site. Raw test scores were converted to age-, sex- and education-corrected z scores using the National Alzheimer's Coordinating Center (NACC) regression-based norming procedures (<https://www.alz.washington.edu/>). Normalized z scores from ten neuropsychological tests were evaluated in determining domain impairment, including three tests per cognitive domain (memory, attention/executive function and language) and one test of global cognition. Impairment in one or more cognitive domains was determined using previously described neuropsychological criteria, and was defined as a score  $>1$  s.d. below norm-referenced values on two or more tests within a single cognitive domain or three or more tests across cognitive domains<sup>36</sup>. Prior studies have established improved sensitivity and specificity of these criteria relative to those employing a single test score, as well as adaptability of this diagnostic approach to various neuropsychological batteries<sup>36,37</sup>. Participants were excluded from cognitive domain analyses if they had less than 90% complete neuropsychological test data (53, 24, and 82 participants were excluded for MRI, PET, and CSF analyses, respectively). Included participants were classified as 0, 1, or 2+ based on the number of cognitive domains for which they had two or more impaired test scores.

Test battery specifics for each UDS version and recruitment site are as follows. i) Global cognition: MMSE for UDS version 2<sup>38</sup> and MoCA for UDS version 3<sup>39</sup>. ii) Memory: The Logical Memory Story A Immediate and Delayed free recall tests (modified from the original Wechsler Memory Scales, Third Edition (WMS-III)) for UDS version 2 and the Craft Stories Immediate and Delayed free recall for UDS version 3. For supplementary tests the USC participants underwent the California Verbal Learning Test, Second Edition (CVLT-II) and the Selective Reminding Test (SRT) sum of free recall trials. Norm-referenced scores for these supplementary test scores were derived from a nationally representative sample published with the test manual (CVLT-II)<sup>40</sup> and in studies of normally ageing adults (SRT). iii) Attention and executive function: The Trails A, Trails B, and Wechsler Adult Intelligence Scale—Revised (WAIS-R) Digit Span Backwards tests for UDS version 2 and the Trails A, Trails B and Digit Span Backwards tests for UDS version 3. iv) Language: The Animal Fluency, Vegetable Fluency, and Boston Naming Tests for UDS version 2 and Animal Fluency, Vegetable Fluency, and Multilingual Naming Test (MINT) for UDS version 3.

## Magnetic resonance imaging and analysis

The MRI data sets were obtained at the Mark and Mary Stevens Neuroimaging and Informatics Institute of USC and Washington University of St. Louis. We developed a standardized high-resolution 3T MRI brain scan protocol. At the USC site, a Siemens 3T Prisma scanner was used with a product 32-channel head receive coil and body transmit coil. At the WashU site, a Siemens 3T mMR with 20-channel head coil and Siemens 3T Vida with 64-channel head coil were used. Anatomical coronal spin echo T2-weighted scans were first obtained through the hippocampi (TR/TE 8020/50 ms, NEX = 1, slice thickness 2 mm with 2 mm gap between slices, FOV = 175 × 175 mm, matrix size = 448 × 448). Baseline coronal T1-weighted maps were then acquired using a T1-weighted 3D volumetric interpolated breath-hold (VIBE) sequence and variable flip angle method using flip angles of 2°, 5°, 10°, 12°, and 15°. Coronal DCE-MRI covering the hippocampi and temporal lobes was acquired using a T1-weighted 3D VIBE sequence (FA = 15°, TR/TE = 5.14/2.18 ms, NEX = 1, slice thickness 5 mm with no gap, FOV 175 × 175 mm, matrix size 320 × 320, voxel size 0.550 × 0.550 × 5 mm<sup>3</sup>). This sequence was repeated for a total of 16 min with an approximate time resolution of 15.4 s. Gadolinium-based contrast agent (GBCA), gadoterate meglumine (Dotarem, Guerbet, France) (0.05 mmol/kg), was administered intravenously into the antecubital vein using a power injector, at a rate of 3 ml/s followed by a 25-ml saline flush, 30 s into the DCE scan.

The standardization and optimization of the MRI protocol required several tests performed on a phantom. Specifically, scanner characterization and calibration sequences including B<sub>0</sub>, T1, and variable flip-angle mapping were implemented, optimized, and applied. After the achievement of good results in terms of quality control and reproducibility, we standardized and employed the same pre-contrast and dynamic T1-weighted protocols at both USC and Washington University sites. Of note, all the other MR sequences were also identical on both scanners.

In order to minimize inter-site variability, the entire MRI protocol including the anatomical and DCE pulse sequences were 100% mirrored from one site to another and the same contrast agent gadoterate meglumine (Dotarem) was injected into participants at the same concentration (0.05 mmol/kg). Finally, exactly the same pre- and post-processing analysis pipeline was applied for both sites, including T1 multi-FA mapping using linear fitting and Patlak-based DCE modeling using the arterial input function determined in each individual from the internal carotid artery. Applying all the above cited factors greatly limited inter-site variability.

The consistency of the results from the two sites was additionally confirmed by our previous publication<sup>7</sup>. In brief, we performed the analysis of the combined DCE data sets from both USC and WashU sites, and additionally site-specific analysis for each of the two sites

separately, which showed no statistically significant differences across sites. Recently, we invited a subset of 52 participants for an additional T1-weighted scan without contrast (using the same scanner and same MR pulse sequences) after their first DCE-MRI<sup>41</sup> and measured both B<sub>0</sub> and T1 values at a two-year interval. This study showed that the results were unchanged and consistent across the scans, supporting minimal intra-site variability.

## Quantification of BBB permeability

See Supplementary Information, Supplementary Methods.

## Quantification of regional brain volumes

HC and PHG morphometry were performed using the FreeSurfer (v5.3.0) software package<sup>42</sup> (<http://surfer.nmr.mgh.harvard.edu/>), as previously described<sup>7</sup>. The HC and PHG were segmented using FreeSurfer Desikan-Killiany and subcortical atlases<sup>43,44</sup>. Then, regional volumes (mm<sup>3</sup>) were derived accordingly. The technical details of this procedure have been described previously<sup>45,46</sup>. Data processing and visualization were performed using the Laboratory of Neuro Imaging (LONI) pipeline system (<http://pipeline.loni.usc.edu>) and Quantitative Imaging Toolkit<sup>47–49</sup>.

## Positron emission tomography and analysis

PET image acquisition was performed at the Molecular Imaging Center of USC or Mallinckrodt Institute of Radiology of WashU. Amyloid and tau PET studies were conducted using <sup>18</sup>F-florbetaben (FBB) or <sup>18</sup>F-florbetapir (FBP) and <sup>18</sup>F-flortaucipir (AV1451), respectively. FBB (Life Molecular Imaging, Inc.) was obtained from SOPHIE, Inc. for the USC site, while FBP was provided by Eli Lilly and Company for the WashU site. For all amyloid PET analysis the FBP and FBB data sets were combined. AV1451 was provided by Avid Radiopharmaceuticals, Inc. for the USC site and was produced by the Mallinckrodt Institute of Radiology for the WashU site. A Siemens Biograph 64 PET scanner was used at the USC site. At the WashU site, FBP scans were acquired on a Siemens mMR and AV1451 scans were acquired on a Siemens Biograph mCT. The mCT session was used for attenuation correction of the mMR scans. Participants were injected with 300 MBq (±10%) of FBB or 370 MBq (±10%) of FBP. FBB and FBP images were acquired from 90 to 110 min and 50 to 70 min, respectively, after injection in accordance with the manufacturers' recommendations. Individuals who participated in amyloid and tau PET studies also had their DCE-MRI scan within 2.2 ± 0.9 and 2.1 ± 0.6 months of their amyloid and tau PET scans, respectively.

In brief, a computed tomography (CT) scan was performed first for attenuation correction before each PET imaging session. The downloaded PET images from FBB, FBP, and AV1451 tracers were processed by using standard uptake value maps (SUV in g/ml). All PET images were co-registered to structural high-resolution 3D T1-weighted Magnetization Prepared Rapid Acquisition Gradient Echo (MP-RAGE) MRI images using FSL-FLIRT (FMRIB's Linear Image Registration Tool)<sup>50</sup>. The FreeSurfer-segmented cerebellum was used as a reference tissue to normalize for both amyloid and tau<sup>51</sup>.

After co-registration of PET images into an anatomical reference image (MNI152 standard-space), Statistical Parametric Mapping (SPM12) was used for group comparison on a voxel-by-voxel basis. Age at time of PET imaging session, sex, and education were introduced in a multiple regression model as covariates. Level of significance was set to  $P < 0.001$  for amyloid and  $P < 0.005$  for tau (uncorrected  $P$  values) with the minimum number of voxels (Ke) in a cluster of 50.

Additionally, given the known AV1451 off-target ligand binding in the choroid plexus (CP)<sup>52,53</sup>, which can contribute to HC regional AV1451 signal owing to the close proximity of the CP to the HC and relatively low spatial resolution of PET scans (that is, ~6-mm voxel size), we took advantage of visualizing CP by DCE-MRI, also performed in these individuals, which allowed us to subtract the contribution of the CP signal to the HC AV1451 proper signal. The following steps were used to correct

for off-target ligand binding to the CP (see Extended Data Fig. 5). Step 1: HC masks were generated from the 3D T1-weighted MP-RAGE. Step 2: CP masks were generated from the T1-weighted VIBE post-GBCA (FA = 15°) image. Step 3: HC and CP masks were overlaid. Step 4: The overlap of the CP and HC masks was subtracted to obtain a CP-corrected HC PET signal after adding 6-mm voxel size on top of the CP mask generated from DCE data. Representative images of HC AV1451 PET signal before and after applying the CP correction are shown in Extended Data Fig. 5.

We next quantified regional changes in amyloid and tau SUV ratio (SUVr) in relation to regional DCE-MRI  $K_{trans}$  values in all participants stratified by *APOE* genotype. The regional SUVr values were taken from the FreeSurfer-segmented HC, PHG, OFC<sup>28</sup>, and ITG<sup>29</sup>. The BBB  $K_{trans}$  constant (DCE-MRI) was determined in all participants (Extended Data Tables 2a, b). This includes those who were analysed for both amyloid and tau ( $n = 58$ ), only amyloid ( $n = 9$ ) or only tau ( $n = 29$ ).

### Lumbar puncture and venipuncture

Participants underwent a lumbar puncture and venipuncture in the morning after an overnight fast. The CSF was collected in polypropylene tubes, processed (centrifuged at 2,000g, 4 °C, 10 min USC site; 5 min WashU site), aliquoted into polypropylene tubes and stored at -80 °C until assay. Blood was collected into EDTA (EDTA) tubes and processed (centrifuged at 2,000g, 4 °C, 10 min USC site; 5 min WashU site). Plasma and buffy coat were aliquoted in polypropylene tubes and stored at -80 °C; buffy coat was used for DNA extraction and *APOE* genotyping.

### *APOE* genotyping

DNA was extracted from buffy coat using the Quick-gDNA Blood Miniprep Kit (catalogue no. D3024, Zymo Research, Irvine, CA). *APOE* genotyping was performed via polymerase chain reaction (PCR)-based retention fragment length polymorphism analysis, as previously reported<sup>7</sup>.

### Molecular assays

**Quantitative western blotting of sPDGFR $\beta$ .** The quantitative western blot analysis was used to detect sPDGFR $\beta$  in human CSF (ng/ml), as previously reported<sup>7,8</sup>.

**BBB breakdown biomarkers.** Albumin quotient ( $Q_{alb}$ , the ratio of CSF to plasma albumin levels) and CSF levels of fibrinogen and plasminogen were determined using enzyme-linked immunosorbent assay (ELISA), as previously reported<sup>7,8</sup>.

**Cyclophilin A.** We developed a novel CypA assay on the Meso Scale Discovery (MSD) platform. Standard-bind 96-well plates (catalogue no. L15XA-3/L11XA-3, MSD, Rockville, MD) were spot-coated with 5  $\mu$ l per well of 40  $\mu$ g/ml rabbit polyclonal anti-CypA antibody (catalogue no. 10436-T52, Sino Biological, Wayne, PA) prepared in 0.03% Triton X-100 in 0.01 M PBS pH 7.4 solution. The plates were left undisturbed overnight to dry at room temperature. The next day, the plates were blocked with 150  $\mu$ l per well of Blocking One (catalogue no. 03953-95, Nacalai Tesque, Japan) and incubated for exactly 1 h with shaking. Meanwhile, samples and standards were prepared in Blocking One blocking buffer. Different concentrations ranging from 3.5 to 200 ng/ml of a recombinant human CypA protein (catalogue no. 3589-CAB, R&D Systems, Minneapolis, MN) were used to generate a standard curve. All CSF samples were diluted 1:3. After blocking, the plates were manually washed three times with 200  $\mu$ l per well of wash buffer (in 0.05% Tween-20 in 0.01 M PBS pH 7.4). The prepared samples or standards were added at 25  $\mu$ l per well, and the plates were incubated overnight at 4 °C with shaking.

The next day, the plates were washed three times, and 25  $\mu$ l per well of 1  $\mu$ g/ml sulfo-tagged mouse monoclonal CypA detection antibody (catalogue no. ab58144, Abcam, Cambridge, MA), prepared in Blocking One. The plates were incubated for 90 min at room temperature

with shaking. Next, the plates were washed four times, then 150  $\mu$ l per well of 2 $\times$  Read Buffer T with surfactant (catalogue no. R92TC-3, MSD, Rockville, MD) was added and the plates were read immediately on an MSD SECTOR Imager 6000 (MSD, Rockville, MD) with electrochemiluminescence detection.

The raw readings were analysed by subtracting the average background value of the zero standard from each recombinant standard and sample reading. A standard curve was constructed by plotting the recombinant standard readings and their known concentrations and applying a nonlinear four-parameter logistics curve fit. The CypA concentrations were calculated using the samples' reading and the standard curve equation; the result was corrected for the sample dilution factor to arrive at the CypA concentration in the CSF samples.

**Matrix metalloproteinase-9.** CSF levels of MMP9 were determined using the human MMP9 Ultra-Sensitive Kit from MSD (cat. No. K151HAC).

**Neuron-specific enolase.** CSF levels of NSE were determined using ELISA (cat. no. E-80NEN, Immunology Consultant Laboratories, Portland, OR). The company no longer sells this product; thus, this analyte was measured in the majority of participants but not in those individuals that enrolled in the study most recently.

**S100B.** CSF levels of the astrocyte-derived cytokine, S100 calcium-binding protein B (S100B), were determined using ELISA (cat. no. EZHS100B-33K, EMD Millipore, Billerica, MA).

**Inflammatory markers.** An MSD multiplex assay was used to determine CSF levels of intercellular adhesion molecule 1 (ICAM1) (cat. no. K15198D, MSD, Rockville, MD), and interleukin-6 (IL6), IL-1 $\beta$ , tumour necrosis factor- $\alpha$  (TNF $\alpha$ ), and interferon gamma (IFN $\gamma$ ) (cat. no. K15049G, MSD, Rockville, MD).

**A $\beta$  peptides.** An MSD multiplex assay (cat. no. K15200E, MSD, Rockville, MD) was used to determine CSF levels of A $\beta$ <sub>1-42</sub>. Participants were stratified based on CSF analysis as either A $\beta$ <sup>+</sup> (<190 pg/ml) or A $\beta$ <sup>-</sup> (>190 pg/ml) using the accepted cutoff values as previously reported for the MSD 6E10 A $\beta$  peptide assay<sup>24</sup>.

**Tau.** Phosphorylated tau (pT181) was determined by ELISA (cat. no. 81581, Innotech, Fujirebio US, Inc., Malvern, PA). Participants were stratified based on CSF analysis as either pTau<sup>+</sup> (>78 pg/ml) or pTau<sup>-</sup> (<78 pg/ml), using the accepted cutoff value as previously reported<sup>25</sup>.

### Human iPSCs

iPSC lines were generated by reprogramming of skin fibroblasts from *APOE*  $\epsilon$ 4/ $\epsilon$ 4 and *APOE*  $\epsilon$ 3/ $\epsilon$ 3 donors with AD or without (control) as recently reported<sup>54</sup>. Reprogramming was performed using integration-free Sendai virus vectors and we passaged cells to passage 15 and confirmed normal karyotype. hiPSCs were maintained on Matrigel (Corning) in mTeSR1 (catalogue no. 85850, StemCell Technologies, Vancouver, BC, Canada) supplemented with 10 ng/ml FGF2 StemBeads (StemCultures) or mTeSR plus (StemCell Technologies) every other day.

### Differentiation of iPSCs into pericytes

Differentiation of iPSCs into pericytes was carried out as described previously<sup>55</sup>. In brief, iPSCs were dissociated with ReLeSR (catalogue no. 05872, StemCell Technologies) and seeded at 55,000 cells/cm<sup>2</sup> in Essential 8 medium (catalogue no. A1517001, ThermoFisher, Waltham, MA, USA) supplemented with ROCK inhibitor Y-27632 (10  $\mu$ M, catalogue no. 72304, StemCell Technologies) on Matrigel (0.5 mg/6-well plate, catalogue no. 354230, Corning, NY, USA). After 24 h incubation, the iPSCs were switched into STEMdiff Mesoderm Induction Medium (MIM, catalogue no. 05221, StemCell Technologies) for 5 days with daily medium change. On day

6 of MIM treatment, the cells were plated on Matrigel at 25,000 cells/cm<sup>2</sup> in pericyte medium (catalogue no. 1201, ScienCell, Carlsbad, CA, USA) for an additional 7 days. The differentiated cells were dissociated with Accutase (catalogue no. 07920, StemCell Technologies). Following incubation with human PDGFR $\beta$  biotinylated antibody (catalogue no. BAF385, R&D Systems, Minneapolis, MN, USA), the cells were incubated with anti-biotin microbeads (catalogue no. 130-090-485, Miltenyi Biotec, Bergisch Gladbach, NRW, Germany) and magnetically sorted using MACS LS columns (catalogue no. 130-042-401, Miltenyi Biotec) following the manufacturer's instructions. Sorted pericytes were plated at a density of 25,000 cells/cm<sup>2</sup> on Matrigel-coated coverslips for immunocytochemistry analyses or poly-L-lysine-coated six-well culture plates for western blot analyses. Differentiated pericytes were positive for the pericyte markers PDGFR $\beta$ , CD13, and NG2, and negative for the endothelial marker CD31, the astrocytic marker GFAP, and the microglial marker CD11b.

## Statistical analyses

Prior to performing statistical analyses, we first screened for outliers using the Grubbs' test, also called the ESD (extreme studentized deviate) method, applying a significance level of  $\alpha = 0.01$  (<https://www.graphpad.com/quickcalcs/grubbs1/>). For each of the outliers identified, a secondary index of outlier influence was applied using the degree of deviation from the mean (greater than  $\pm 3$  s.d.)<sup>56</sup>. Using these stringent criteria, a total of five outliers (one each in Figs. 1j, k and 2j and one each in Extended Data Fig. 6a, b, were removed from analyses, as indicated in the legends of these figures. Continuous variables were also evaluated for departures from normality through quantitative examination of skewness and kurtosis, in addition to visual inspection of frequency distributions. Where departures of normality were identified, log<sub>10</sub> transformations were applied, and distribution normalization was confirmed before parametric analyses. This was done for Fig. 4h, k and Extended Data Fig. 7a, b, d, e. As the use of log<sub>10</sub> transformations accounts for any non-normality, this obviated the need for outliers exclusion.

**DCE-MRI  $K_{trans}$  and CSF sPDGFR $\beta$  and CypA.** Regional DCE-MRI  $K_{trans}$  values and CSF sPDGFR $\beta$ , CypA and MMP9 levels were compared across the entire sample stratified by *APOE* status. As in the *APOE4* group relatively few participants were homozygous  $\epsilon 4/\epsilon 4$  compared to heterozygous  $\epsilon 3/\epsilon 4$  (14% for DCE-MRI analysis, and 18% for sPDGFR $\beta$  analysis), and initial comparisons between  $\epsilon 4/\epsilon 4$  and  $\epsilon 3/\epsilon 4$  carriers did not show any significant differences in regional HC and PHG DCE-MRI  $K_{trans}$  values (CDR 0,  $P_{HC} = 0.19$  and  $P_{PHG} = 0.54$  (PHG); CDR 0.5,  $P_{HC} = 0.22$  and  $P_{PHG} = 0.84$ ) or CSF sPDGFR $\beta$  levels (CDR 0,  $P = 0.23$ ; CDR 0.5,  $P = 0.47$ ), all subsequent analyses combined *APOE4* carriers ( $\epsilon 3/\epsilon 4$  and  $\epsilon 4/\epsilon 4$ ), and compared these participants to *APOE3* carriers ( $\epsilon 3/\epsilon 3$ ) stratified by cognitive impairment status (CDR 0 versus 0.5 and 0 versus 1 versus 2+ cognitive domain impairment using ANCOVA with FDR correction for multiple comparisons (see details below). For CDR analyses, model covariates included age, sex, and education. Cognitive domain impairment was determined using age-, sex-, and education-corrected values, so these covariates were not additionally included in the analyses. Additional post hoc ANCOVA analyses evaluated whether the observed differences remained significant after stratifying *APOE4* carriers by CSF A $\beta_{1-42}$  and pTau status, and after statistically controlling for CSF A $\beta_{1-42}$  and pTau status and regional brain volume in *APOE4* non-carriers and carriers. These findings were also confirmed by hierarchical logistic regression models using the same covariates.

**PET AD biomarkers.** In a subset of participants who underwent amyloid and tau PET imaging together with DCE-MRI studies, we used ANCOVA models controlled for age, sex and education to compare regional amyloid and tau ligand binding and DCE-MRI values in a set of *APOE4* non-carriers and carriers within a priori regions of interest, based on prior imaging studies, to determine whether distinct regional pathologies differed by *APOE4* carrier status.

**Baseline CSF sPDGFR $\beta$  as a continuous predictor of cognitive decline.** For linear mixed model analysis, baseline CSF sPDGFR $\beta$  was a continuous predictor of demographically corrected global cognitive change at 2-year follow up intervals, controlling for CSF A $\beta_{1-42}$  and CSF pTau status. Global cognition was indexed by age-, sex-, and education-corrected z scores on mental status exam (MMSE or MoCA) and as the global cognitive composite of all age-, sex-, and education-corrected neuropsychological test z scores (see above for list of neuropsychological tests). Time was modelled with date of LP as baseline (t0) with two follow-up intervals of 2 years each (t1, t2). Additional analyses confirmed all findings when time was modelled as time since baseline, with date of lumbar puncture as baseline (t0) and follow up as annual intervals (t1–n).

All longitudinal mixed models treated CSF sPDGFR $\beta$  as a continuous predictor. Although we have previously established that CSF sPDGFR $\beta$  is a marker of pericyte injury<sup>7,8,57</sup>, the optimal cutoff value for abnormal CSF sPDGFR $\beta$  levels indicative of pericyte injury remains unknown. Autopsy studies are required to determine optimal in vivo biomarker cutoff values that predict gold-standard neuropathological measures, such as studies conducted for CSF and PET markers of amyloid and tau. Given the lack of available autopsy data relating CSF sPDGFR $\beta$  to neuropathological markers of pericyte injury, we chose to divide participants by CSF sPDGFR $\beta$  values using median split for the purposes of visual display only (higher CSF sPDGFR $\beta$  was above sample median and lower CSF sPDGFR $\beta$  was below sample median). The median split was not used in statistical analyses and was only used for the purpose of visual display (Fig. 3a) for statistical parameters from analyses using CSF sPDGFR $\beta$  as a continuous predictor of cognitive decline).

**Correlational analyses.** Pearson product moment correlations were used to evaluate relationships among CSF sPDGFR $\beta$ , CypA, MMP9, fibrinogen, plasminogen and hippocampal and parahippocampal BBB  $K_{trans}$  levels among *APOE4* carriers.

**Multiple comparison correction and missing data.** Given the large number of analyses, FDR correction was applied to *P* values for primary study outcomes (DCE-MRI, sPDGFR $\beta$ ) evaluated in the entire sample by *APOE4* carrier status and CDR status using the Benjamini–Hochberg method<sup>58</sup> in ANCOVA and logistic regression models controlling for age, sex, education, brain volume, and CSF A $\beta_{1-42}$  and pTau status (for DCE-MRI analyses). Post hoc confirmatory analyses in participant subsets further evaluating independence of CSF and PET markers of amyloid and tau, evaluation of mechanistic markers (that is, CypA and MMP9), and longitudinal analysis of predictive value of CSF sPDGFR $\beta$  were not corrected for multiple comparisons. For longitudinal data with variable follow up, we used linear mixed model analyses with and accounted for missing data via the missing at random assumption.

## Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

## Data availability

All data generated and/or analysed during this study are either included in this article (and its Supplementary Information) or are available from the corresponding author on reasonable request. Source Data for Figs. 1–4 are provided with the article.

## Code availability

All software used in this study are publicly available: Rocketship v1.2 ([https://github.com/petmri/ROCKETSHIP/blob/master/dce/com- pare\\_gui.m](https://github.com/petmri/ROCKETSHIP/blob/master/dce/com- pare_gui.m)), FreeSurfer (v5.3.0) (<http://surfer.nmr.mgh.harvard.edu/>),



FSL-FLIRT (<https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FLIRT>), SPM12 (<https://www.fil.ion.ucl.ac.uk/spm/software/spm12/>), and Quantitative Imaging Toolkit (<https://cabeen.io/about/publication/cabeen2018quantitative/>).

32. Morris, J. C. et al. The Uniform Data Set (UDS): clinical and cognitive variables and descriptive data from Alzheimer Disease Centers. *Alzheimer Dis. Assoc. Disord.* **20**, 210–216 (2006).
33. Morris, J. C. The Clinical Dementia Rating (CDR): current version and scoring rules. *Neurology* **43**, 2412–2414 (1993).
34. Nation, D. A. et al. Antemortem pulse pressure elevation predicts cerebrovascular disease in autopsy-confirmed Alzheimer's disease. *J. Alzheimers Dis.* **30**, 595–603 (2012).
35. Bangen, K. J. et al. Aggregate effects of vascular risk factors on cerebrovascular changes in autopsy-confirmed Alzheimer's disease. *Alzheimers Dement.* **11**, 394–403.e1 (2015).
36. Jak, A. J. et al. Quantification of five neuropsychological approaches to defining mild cognitive impairment. *Am. J. Geriatr. Psychiatry* **17**, 368–375 (2009).
37. Jak, A. J. et al. Neuropsychological criteria for mild cognitive impairment and dementia risk in the Framingham heart study. *J. Int. Neuropsychol. Soc.* **22**, 937–943 (2016).
38. Weintraub, S. et al. The Alzheimer's Disease Centers' Uniform Data Set (UDS): the neuropsychologic test battery. *Alzheimer Dis. Assoc. Disord.* **23**, 91–101 (2009).
39. Besser, L. et al. Version 3 of the National Alzheimer's Coordinating Center's Uniform Data Set. *Alzheimer Dis. Assoc. Disord.* **32**, 351–358 (2018).
40. Delis, D., Kramer, J., Kaplan, E. & Ober, B. *California Verbal Learning Test* (PsychCorp, 2000).
41. Montagne, A. et al. Undetectable gadolinium brain retention in individuals with an age-dependent blood-brain barrier breakdown in the hippocampus and mild cognitive impairment. *Alzheimers Dement.* **15**, 1568–1575 (2019).
42. Fischl, B. FreeSurfer. *Neuroimage* **62**, 774–781 (2012).
43. Fischl, B. et al. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron* **33**, 341–355 (2002).
44. Desikan, R. S. et al. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage* **31**, 968–980 (2006).
45. Fischl, B. & Dale, A. M. Measuring the thickness of the human cerebral cortex from magnetic resonance images. *Proc. Natl Acad. Sci. USA* **97**, 11050–11055 (2000).
46. Fischl, B., Sereno, M. I., Tootell, R. B. & Dale, A. M. High-resolution intersubject averaging and a coordinate system for the cortical surface. *Hum. Brain Mapp.* **8**, 272–284 (1999).
47. Dinov, I. et al. Neuroimaging study designs, computational analyses and data provenance using the LONI pipeline. *PLoS ONE* **5**, e13070 (2010).
48. Sepehrband, F. et al. Neuroanatomical morphometric characterization of sex differences in youth using statistical learning. *Neuroimage* **172**, 217–227 (2018).
49. Cabeen, R. P., Laidlaw, D. H. & Toga, A. W. Quantitative imaging toolkit: software for interactive 3D visualization, data exploration, and computational analysis of neuroimaging datasets. *Proc. Intl Soc. Magnetic Resonance in Medicine (ISMRM)* vol. **2854** (2018).
50. Jenkinson, M., Bannister, P., Brady, M. & Smith, S. Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage* **17**, 825–841 (2002).
51. Bullich, S. et al. Optimal reference region to measure longitudinal amyloid- $\beta$  change with  $^{18}\text{F}$ -florbetaben PET. *J. Nucl. Med.* **58**, 1300–1306 (2017).
52. Marquié, M. et al. Lessons learned about [F-18]-AV-1451 off-target binding from an autopsy-confirmed Parkinson's case. *Acta Neuropathol. Commun.* **5**, 75 (2017).
53. Mishra, S. et al. AV-1451 PET imaging of tau pathology in preclinical Alzheimer disease: defining a summary measure. *Neuroimage* **161**, 171–178 (2017).
54. TCW, J. et al. Cholesterol and matrisome pathways dysregulated in human  $\epsilon 4$  glia. Preprint at <https://www.biorxiv.org/content/10.1101/713362v1> (2019).
55. Faal, T. et al. Induction of mesoderm and neural crest-derived pericytes from human pluripotent stem cells to study blood-brain barrier interactions. *Stem Cell Reports* **12**, 451–460 (2019).
56. Aggarwal, C. C. *Outlier Analysis* (Springer, 2013).
57. Sagare, A. P., Sweeney, M. D., Makshanoff, J. & Zlokovic, B. V. Shedding of soluble platelet-derived growth factor receptor- $\beta$  from human brain pericytes. *Neurosci. Lett.* **607**, 97–101 (2015).
58. Glickman, M. E., Rao, S. R. & Schultz, M. R. False discovery rate control is a recommended alternative to Bonferroni-type adjustments in health studies. *J. Clin. Epidemiol.* **67**, 850–857 (2014).

**Acknowledgements** The work of B.V.Z. is supported by National Institutes of Health (NIH) grant nos. R01AG023084, R01NS034467, R01AG039452, 5P01AG052350 and 5P50AG005142, in addition to an Alzheimer's Association strategic 509279 grant, Cure Alzheimer's Fund, the Foundation Leducq Transatlantic Network of Excellence for the Study of Perivascular Spaces in Small Vessel Disease reference no. 16 CVD 05, and Open Philanthropy. D.A.N. is supported by NIH grant nos. R01AG060049, R01AG064228, P01AG052350 and P50AG016573, Alzheimer's Association grant AARG-17-532905 and Alzheimer's Association strategic grant 509279. D.P.B. and M.G.H. are supported by the L. K. Whittier Foundation, grant nos. P01AG052350, R01AG054434 and R01AG055770. Enrolment of participants into the WashU Knight ADRC is supported by NIH grant nos. P50AG05681, P01AG03991 and P01AG026276 (J.C.M.). E.M.R. is supported by National Institute of Aging (NIA) grant nos. P30AG19610 and R01AG031581, in addition to the state of Arizona. Enrolment of participants into the USC ADRC is supported by NIH grant no. 5P50AG005142 (H.C.C.). Avid Radiopharmaceuticals, a wholly owned subsidiary of Eli Lilly and Company, provided doses of FBP and financial support for FBP scanning at the WashU site. Avid Radiopharmaceuticals also provided the WashU site with AV1451 precursor and technology transfer for producing the tracer on site. Avid Radiopharmaceuticals was not involved in data analysis or interpretation.

**Author contributions** A.M., D.A.N., A.P.S., G.B., M.D.S. and B.V.Z. designed the research study and analysed and interpreted the data. A.M., D.A.N., A.P.S., G.B., M.D.S., A.C., M.P. and Y.C. performed the experiments and analysed the data. A.M. and G.B. performed the MRI analysis. A.M., G.B. and A.C. performed the PET analysis. A.P.S., M.D.S. and M.P. performed the biofluids analysis. D.A.N. performed the neuropsychological analysis. A.P.S., Y.C., B.V.Z. and J.TCW. contributed to human iPSC-pericyte experiments. L.M.D. and A.R.N. prepared and submitted the study to the IRB. M.P., E.J., D.P.B., M.G.H., T.L.S.B., A.M.F., J.M.R., L.S.S., J.C.M., E.M.R., R.J.C., H.C.C., J.TCW., J.P., P.S.C., M.L. and A.W.T. recruited the participants and performed and provided the imaging scans. A.P.S., G.B., M.G.H., T.L.S.B., A.M.F., J.M.R., L.S.S., J.C.M., E.M.R., R.J.C., H.C.C., J.TCW., P.S.C. and A.W.T. provided critical reading of the manuscript. A.M. and D.A.N. contributed to manuscript writing and B.V.Z. wrote the manuscript.

**Competing interests** The authors declare no competing interests.

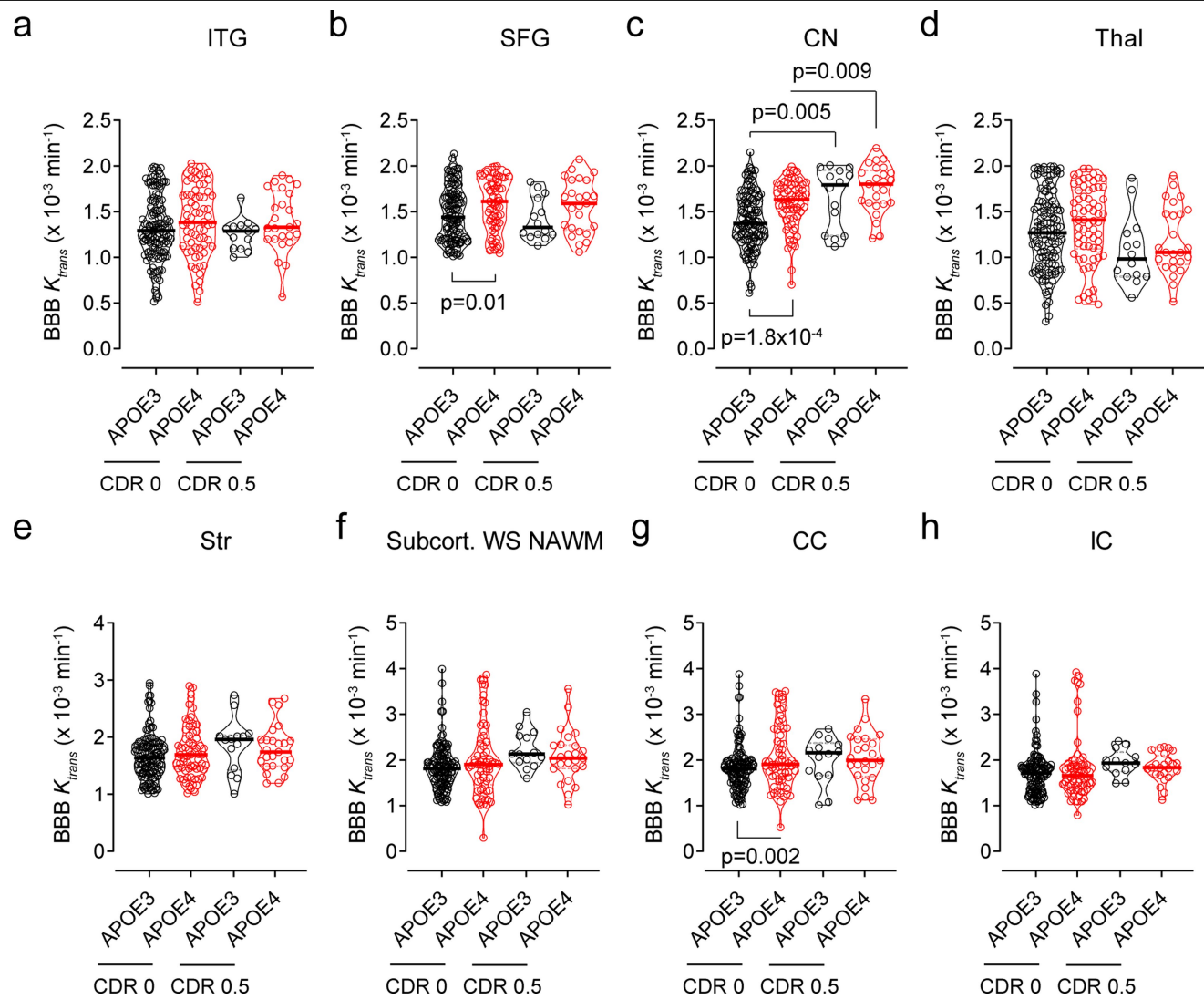
#### Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41586-020-2247-3>.

**Correspondence and requests for materials** should be addressed to B.V.Z.

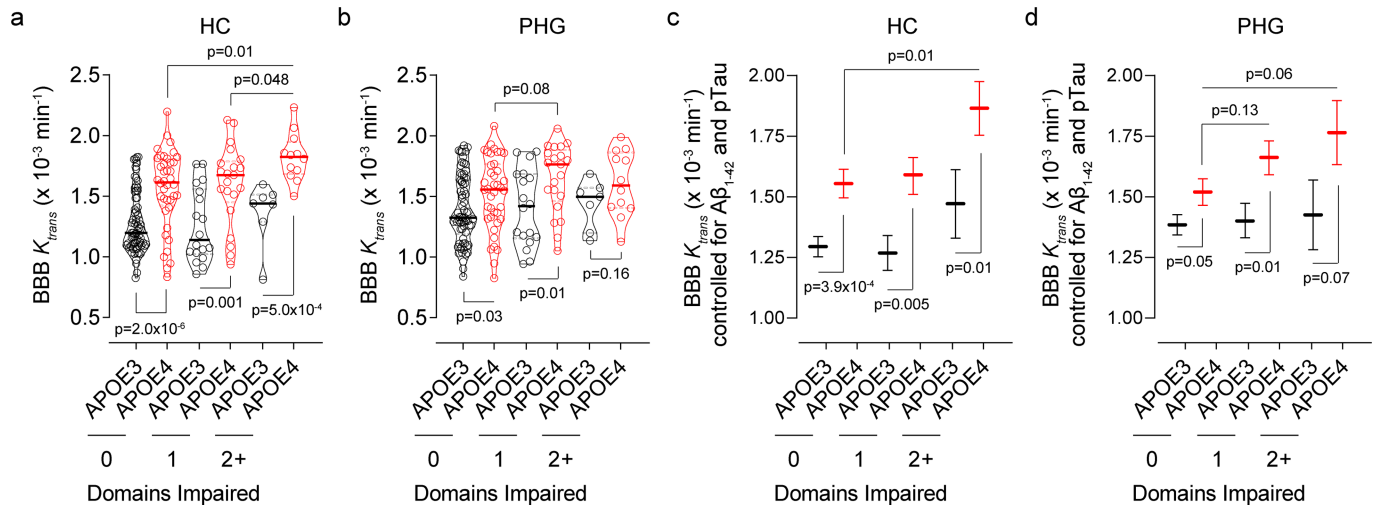
**Peer review information** *Nature* thanks Ronald Thomas, Yi-Fen Yen and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>.



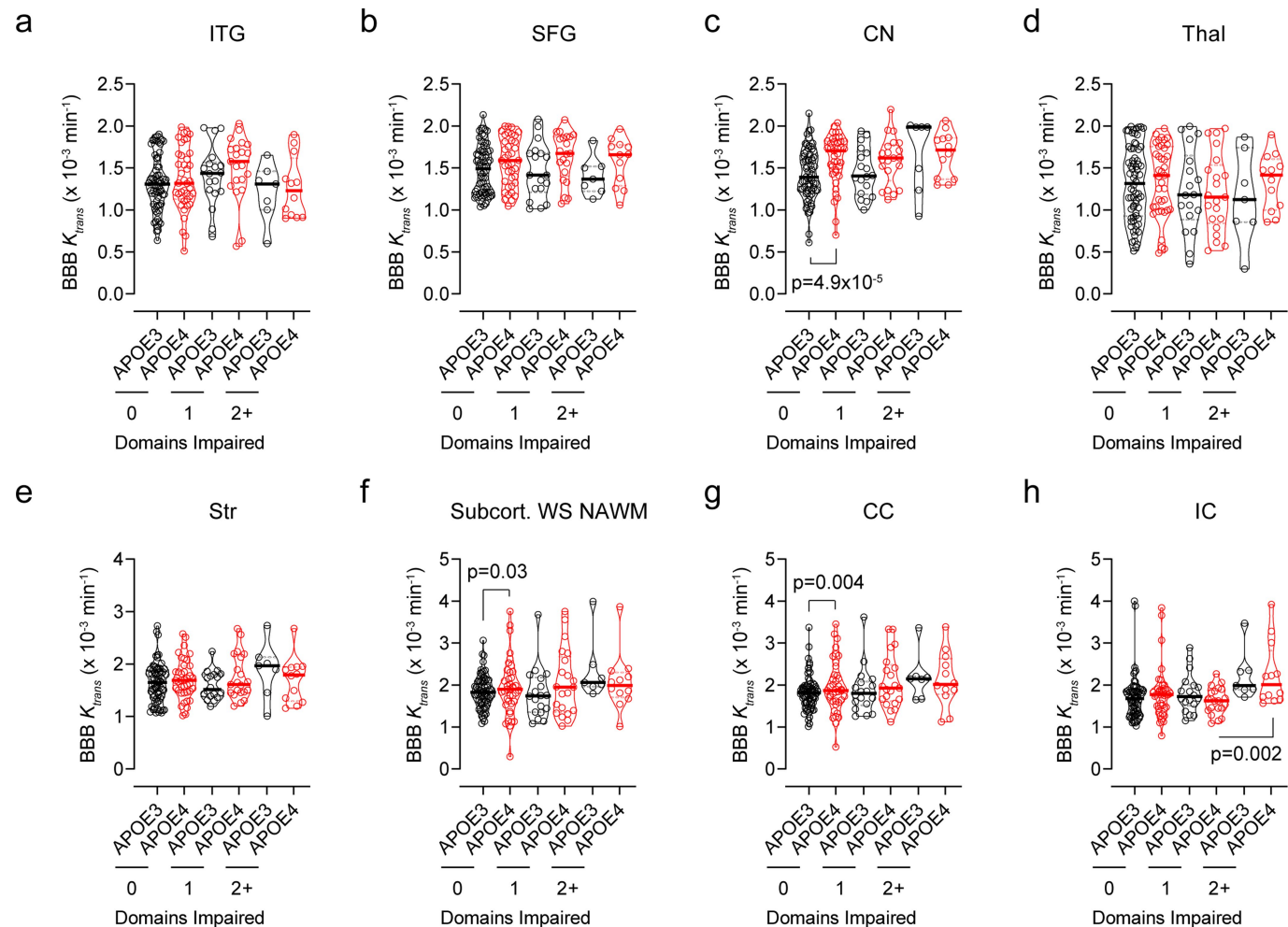
**Extended Data Fig. 1 | Regional BBB  $K_{trans}$  constant in eight additional brain regions in *APOE4* carriers and non-carriers (*APOE3*) with CDR status 0 and 0.5.** BBB  $K_{trans}$  constant in the ITG (**a**), superior frontal gyrus (SFG, **b**), caudate nucleus (CN, **c**), thalamus (Thal, **d**), striatum (Str, **e**), subcortical watershed normal-appearing white matter (Subcort. WS NAWM, **f**), corpus callosum (CC, **g**),

and internal capsule (IC, **h**) in individuals with CDR 0 bearing *APOE3* (black,  $n=128$ ) and *APOE4* (red,  $n=68$ ), and with CDR 0.5 bearing *APOE3* (black,  $n=14$ ) and *APOE4* (red,  $n=25$ ). Continuous lines, median; dotted lines, IQR. Significance by ANCOVAs for main effects and post hoc comparisons controlling for age, sex, and education.



**Extended Data Fig. 2 | BBB breakdown in the HC and PHG in *APOE4* carriers increases with cognitive domain impairment. a, b,  $K_{trans}$  constant in the HC (a) and PHG (b) in individuals with no cognitive domains impaired bearing *APOE3* (black,  $n=70$ ) or *APOE4* (red,  $n=40$ ); one cognitive domain impaired bearing *APOE3* ( $n=18$ ) or *APOE4* ( $n=21$ ); and two or more cognitive domains impaired bearing *APOE3* ( $n=7$ ) or *APOE4* ( $n=12$ ). Continuous lines, median; dotted lines, IQR. c, d,  $K_{trans}$  (estimated marginal mean  $\pm$  s.e.m. from ANCOVA models corrected for age, sex, education, CSF  $A\beta_{1-42}$  and pTau status, and HC**

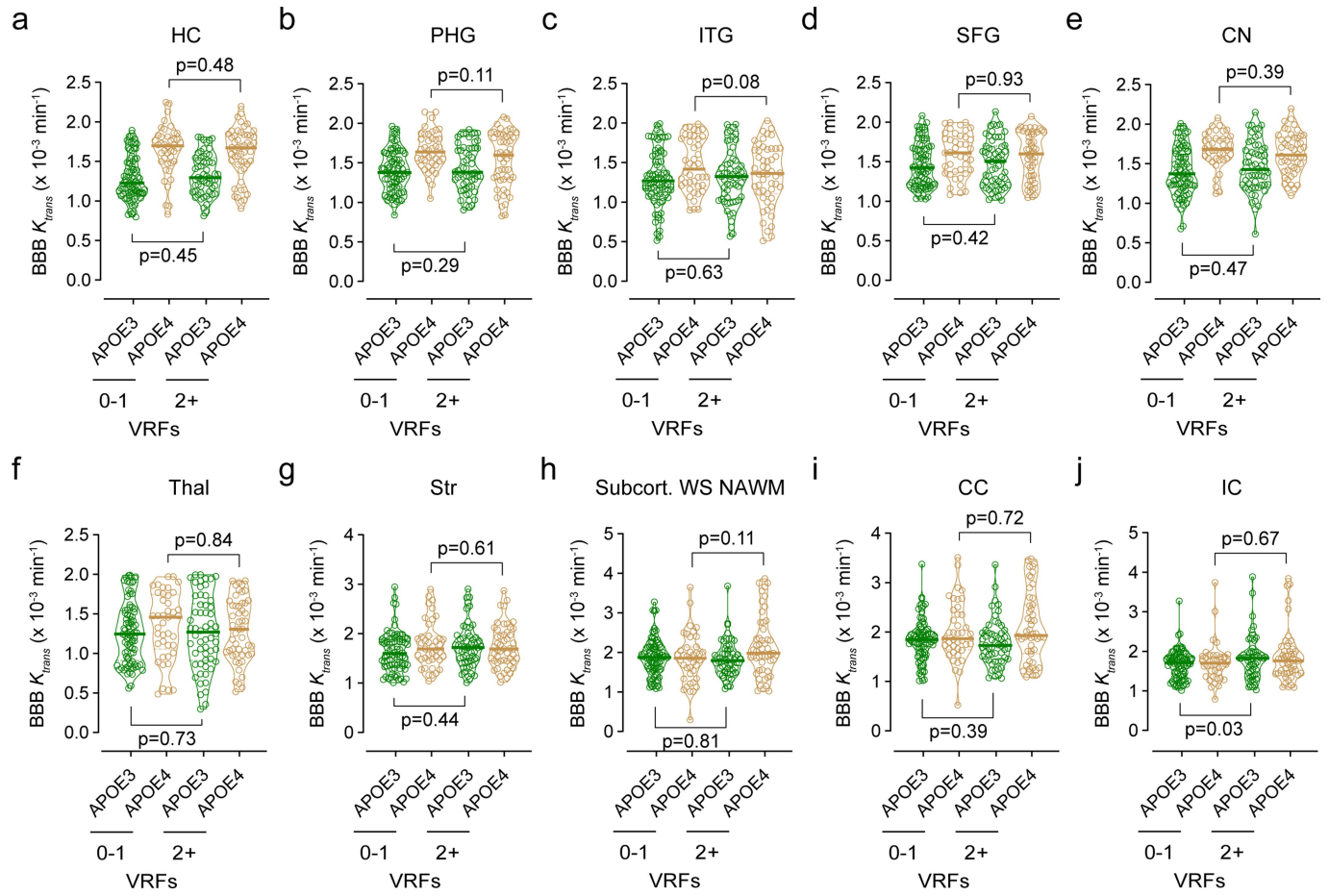
and PHG volumes) in the HC (c) and PHG (d) in individuals with no cognitive domains impaired bearing *APOE3* ( $n=70$ ) or *APOE4* ( $n=40$ ); one cognitive domain impaired bearing *APOE3* ( $n=18$ ) or *APOE4* ( $n=21$ ); and two or more cognitive domains impaired bearing *APOE3* ( $n=7$ ) or *APOE4* ( $n=12$ ). Significance by ANCOVA for main effects and post hoc comparisons controlling for age, sex, and education. All ANCOVA omnibus tests remained significant at FDR threshold of 0.05.



**Extended Data Fig. 3 | Regional BBB  $K_{trans}$  constant in eight additional brain regions in *APOE4* carriers and *APOE3* carriers with different degrees of cognitive domain impairment.**  $K_{trans}$  constant in the ITG (a), SFG (b), CN (c), thalamus (d), striatum (e), subcortical WS NAWM (f), CC (g), and IC (h) in individuals with no cognitive domains impaired bearing *APOE3* (black,  $n = 70$ )

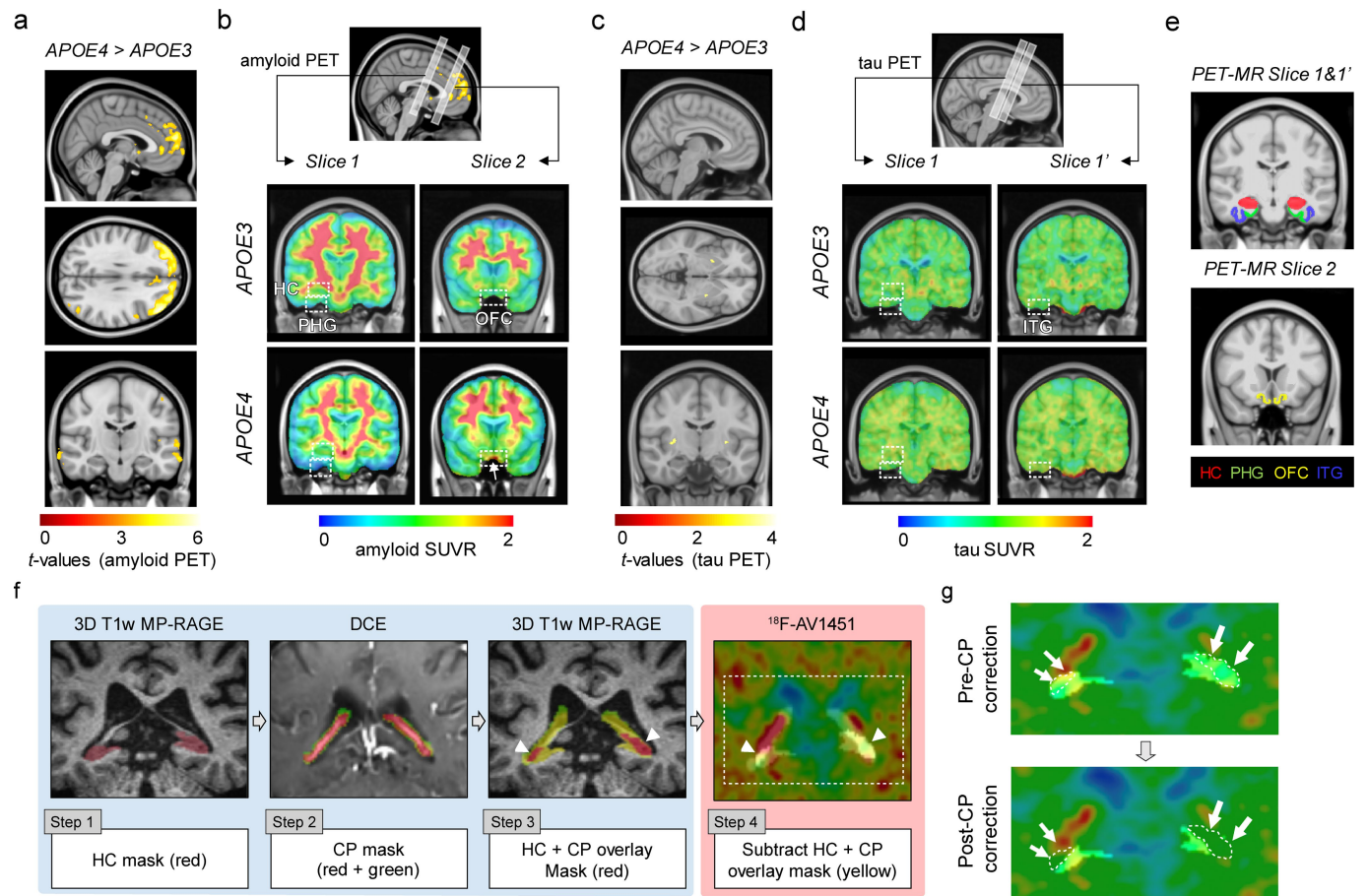
or *APOE4* (red,  $n = 40$ ): one cognitive domain impaired bearing *APOE3* ( $n = 18$ ) or *APOE4* ( $n = 21$ ); and two or more cognitive domains impaired bearing *APOE3* ( $n = 7$ ) or *APOE4* ( $n = 12$ ). Continuous lines, median; dotted lines, IQR. Significance tests from ANCOVAs for main effects and post hoc comparisons controlling for age, sex, and education.





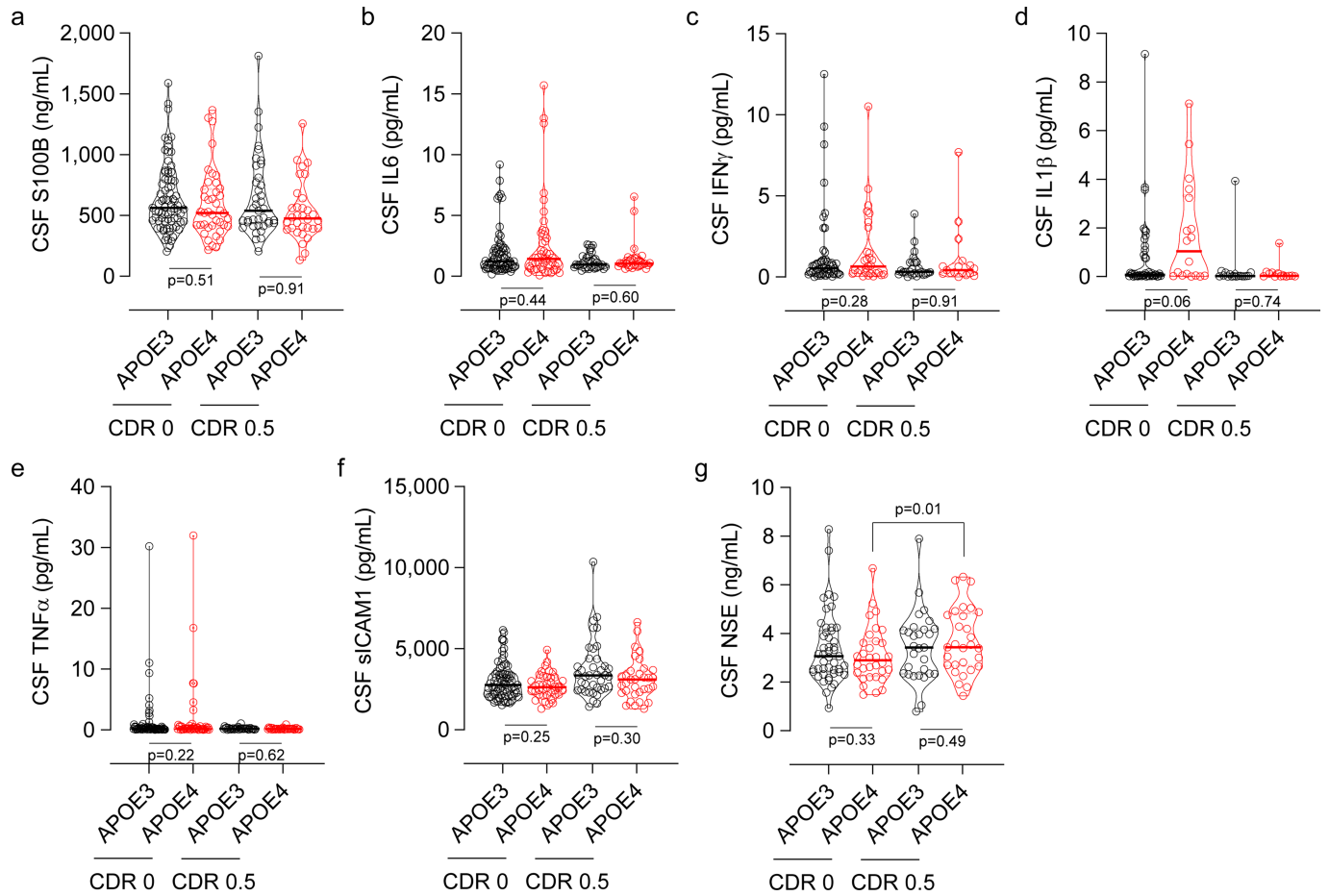
**Extended Data Fig. 4 | Regional BBB  $K_{trans}$  constant in all studied brain regions in *APOE4* carriers and *APOE3* carriers in relation to vascular risk factors.**  $K_{trans}$  constant in the HC (a), PHG (b), ITG (c), SFG (d), CN (e), thalamus (f), striatum (g), subcortical WS NAWM (h), CC (i), and IC (j) in *APOE3* (green,

$n = 80$ ) and *APOE4* (brown,  $n = 42$ ) carriers with 0–1 vascular risk factors (VRFs), and *APOE3* ( $n = 58$ ) and *APOE4* ( $n = 51$ ) carriers with 2+ VRFs. Continuous lines, medians; dotted lines, IQR. Significance by ANCOVAs for main effects and post hoc comparisons controlling for age, sex, and education.



**Extended Data Fig. 5 | Amyloid and tau PET analysis in  $APOE4$  carriers and correction of  $^{18}F$ -AV1451 off-target binding in the choroid plexus.** All studies were performed in individuals with CDR score 0. Amyloid and tau PET studies were conducted using  $^{18}F$ -FBB or  $^{18}F$ -FBP, and  $^{18}F$ -AV1451, respectively. For amyloid PET data analysis, FBP and FBB data sets were combined. **a**, Uptake of amyloid tracers by the OFC in  $APOE4$  ( $n = 29$ ) relative to  $APOE3$  ( $n = 45$ ) carriers (voxel-wise two-sample one-tailed  $t$ -tests). **b**, Representative amyloid PET SUVR maps from an  $APOE3$  homozygote (top) and an  $APOE4$  carrier ( $APOE4$ ) (bottom). Slices 1 and 2, regions of interest (ROIs) for amyloid PET and BBB DCE-MRI scans (see **e**). Arrow, amyloid tracer uptake by OFC. The  $APOE3$  and  $APOE4$  representative images used FBP. **c**, Uptake of tau tracer shows undetectable tau accumulation in  $APOE3$  ( $n = 60$ ) or  $APOE4$  ( $n = 37$ ) carriers (voxel-wise two-sample one-tailed  $t$ -tests). **d**, Representative tau PET SUVR

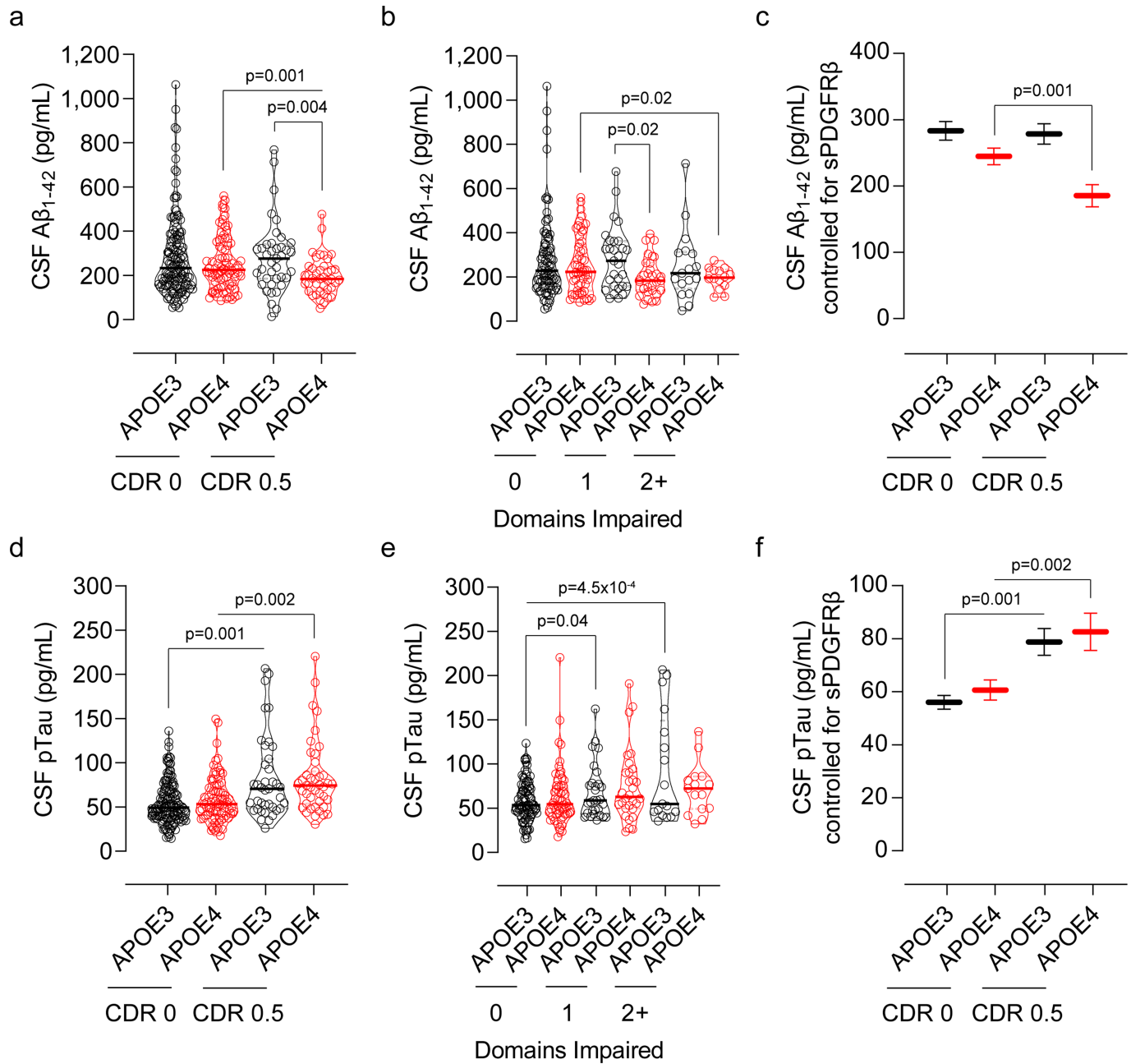
maps from  $APOE3$  (top) and  $APOE4$  (bottom) carriers. Slices 1 and 1', ROIs for tau PET and BBB DCE-MRI scans, respectively (see **e**). **e**, Coronal 3D scans of regions studied in Fig. 2: HC (red), PHG (green), medial OFC (yellow), and ITG (blue). **f**, Correction of  $^{18}F$ -AV1451 off-target binding in the choroid plexus. Step 1, HC masks were generated from the 3D T1-weighted magnetization prepared-rapid gradient echo (MP-RAGE). Step 2, CP masks were generated from the T1-weighted VIBE post-GBCA image (flip angle,  $15^\circ$ ). Step 3, HC and CP masks were overlaid (arrowheads, red). Step 4, areas of CP overlap with HC masks (arrowheads, yellow) were subtracted to obtain CP-corrected HC tau PET signal after adding 6-mm voxel size on top of CP mask generated from DCE data. **g**, Representative images of HC tau PET signal before (top) and after (bottom) applying the CP correction (arrows and white dotted lines show overlap between HC and CP).



**Extended Data Fig. 6 | CSF biomarkers of glia and inflammatory response and endothelial and neuronal cell injury in *APOE4* and *APOE3* carriers.**

**a**, CSF astrocytic S100B levels in individuals with CDR 0 bearing *APOE3* (black,  $n=77$ ) or *APOE4* (red,  $n=41$ ), and with CDR 0.5 bearing *APOE3* ( $n=39$ ) or *APOE4* ( $n=32$ ). **b**, CSF IL6 levels in individuals with CDR 0 bearing *APOE3* ( $n=71$ ) or *APOE4* ( $n=47$ ), and with CDR 0.5 bearing *APOE3* ( $n=34$ ) or *APOE4* ( $n=32$ ). **c**, CSF IFN $\gamma$  levels in individuals with CDR 0 bearing *APOE3* ( $n=54$ ) or *APOE4* ( $n=29$ ), and with CDR 0.5 bearing *APOE3* ( $n=25$ ) or *APOE4* ( $n=17$ ). **d**, CSF IL1 $\beta$  levels in individuals with CDR 0 bearing *APOE3* ( $n=43$ ) or *APOE4* ( $n=18$ ), and with CDR 0.5 bearing *APOE3* ( $n=17$ ) or *APOE4* ( $n=13$ ). **e** CSF TNF $\alpha$  levels in individuals

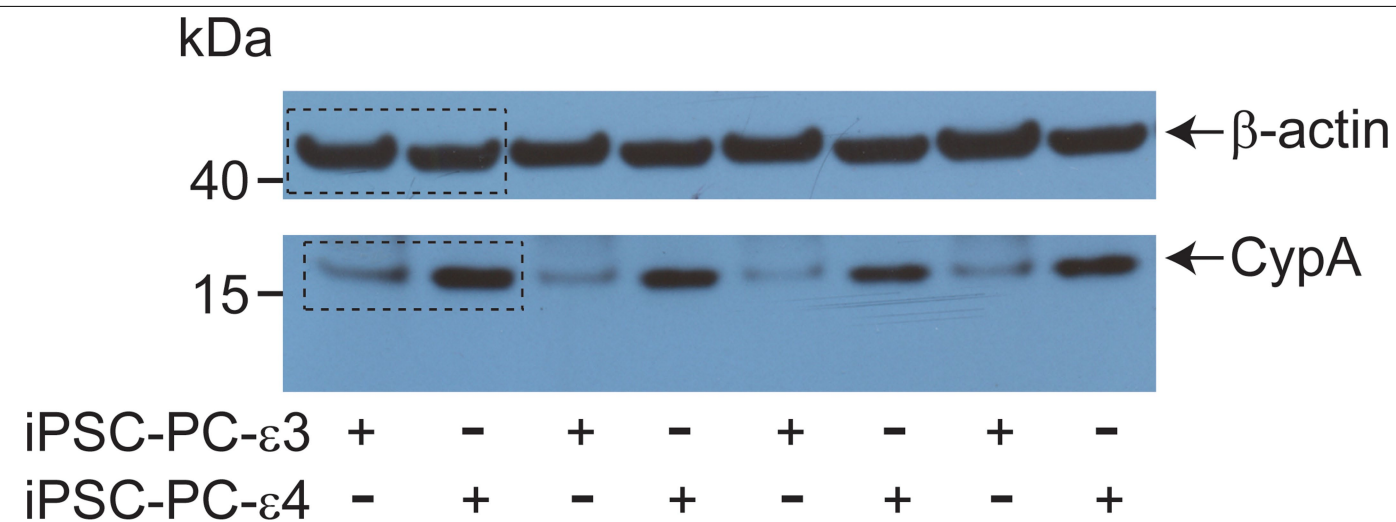
with CDR 0 bearing *APOE3* ( $n=70$ ) or *APOE4* ( $n=46$ ), and with CDR 0.5 bearing *APOE3* ( $n=34$ ) or *APOE4* ( $n=32$ ). **f**, CSF soluble intercellular adhesion molecule 1 (sICAM1) levels in individuals with CDR 0 bearing *APOE3* ( $n=77$ ) or *APOE4* ( $n=40$ ), and with CDR 0.5 bearing *APOE3* ( $n=39$ ) or *APOE4* ( $n=33$ ). **g**, CSF NSE levels in individuals with CDR 0 bearing *APOE3* ( $n=47$ ) or *APOE4* ( $n=32$ ), and with CDR 0.5 bearing *APOE3* ( $n=29$ ) or *APOE4* ( $n=29$ ). Continuous lines, median; dotted lines, IQR. **a** and **b** had one outlier each, which were removed before statistical analysis (see Methods). Significance by ANCOVAs for main effects and post hoc comparisons controlling for age, sex, and education.



**Extended Data Fig. 7 | Decreased CSF Aβ<sub>1-42</sub> and increased pTau levels in APOE4 carriers with cognitive impairment.** **a**, CSF Aβ<sub>1-42</sub> levels in individuals with CDR 0 bearing APOE3 (black,  $n=141$ ) or APOE4 (red,  $n=83$ ) and with CDR 0.5 bearing APOE3 ( $n=39$ ) or APOE4 ( $n=41$ ). **b**, CSF Aβ<sub>1-42</sub> levels in APOE3 ( $n=89$ ) and APOE4 ( $n=55$ ) carriers with no cognitive domains impaired, APOE3 ( $n=29$ ) and APOE4 ( $n=31$ ) carriers with one cognitive domain impaired, and APOE3 ( $n=17$ ) and APOE4 ( $n=14$ ) carriers with two or more cognitive domains impaired. **c**, CSF Aβ<sub>1-42</sub> levels (estimated marginal means  $\pm$  s.e.m. from ANCOVA models corrected for age, sex, education, and CSF sPDGFRβ levels) in individuals with CDR 0 bearing APOE3 ( $n=141$ ) or APOE4 ( $n=83$ ) and with CDR 0.5 bearing APOE3 ( $n=39$ ) or APOE4 ( $n=41$ ). **d**, CSF pTau levels in individuals with CDR 0 bearing APOE3 ( $n=141$ ) or APOE4 ( $n=82$ ) and with CDR 0.5 bearing

APOE3 ( $n=39$ ) or APOE4 ( $n=43$ ). **e**, CSF pTau levels in APOE3 ( $n=89$ ) and APOE4 ( $n=56$ ) carriers with no cognitive domains impaired, APOE3 ( $n=29$ ) and APOE4 ( $n=30$ ) carriers with one cognitive domain impaired, and APOE3 ( $n=17$ ) and APOE4 ( $n=15$ ) carriers with two or more cognitive domains impaired. **f**, CSF pTau levels (estimated marginal means  $\pm$  s.e.m. from ANCOVA models corrected for age, sex, education, and CSF sPDGFRβ levels) in individuals with CDR 0 bearing APOE3 ( $n=141$ ) or APOE4 ( $n=82$ ) and with CDR 0.5 bearing APOE3 ( $n=39$ ) or APOE4 ( $n=43$ ). Violin plots: continuous lines, median; dotted lines, IQR. CSF Aβ<sub>1-42</sub> and pTau values were log<sub>10</sub>-transformed before statistical analysis because they had a non-normal distribution. Significance tests from ANCOVAs for main effects and post hoc comparisons controlling for age, sex, and education.





**Extended Data Fig. 8 | Full scans of western blots.** Full scans of western blots for CypA shown in Fig. 4m (top).

Extended Data Table 1 | *APOE3* and *APOE4* carriers studied for regional BBB permeability changes by DCE-MRI

APOE Genotype	<i>APOE3</i>	<i>APOE4</i>	<i>APOE3</i>	<i>APOE4</i>
Clinical Dementia Rating (CDR)	0	0	0.5	0.5
No. of participants	130	76	14	25
Age at MRI, years, Mean (SD)	69.9 (7.9)	67.3 (8.7)	73.8 (8.3)	69.4 (8.7)
Female, %	62.3	57.9	42.9	56
Education, years, Mean (SD)	16.6 (2.7)	16.7 (2.0)	16.4 (2.5)	17.1 (2.1)
Cognitive domain impairment, No. 0, 1, 2+	78, 17, 2	38, 16, 4	6, 2, 6	6, 8, 9
Vascular risk factors, No. 0-1, 2+	77, 53	38, 38	8, 6	10, 15

Extended Data Table 2 | *APOE3* and *APOE4* carriers studied for regional amyloid or tau brain accumulation by PET and BBB permeability changes by DCE-MRI

<b>APOE Genotype</b>	<b><i>APOE3</i></b>	<b><i>APOE4</i></b>
<b>Clinical Dementia Rating (CDR)</b>	0	0
<b>No. of participants</b>	45	29
<b>No. of participants (FBB, FBP)</b>	5, 40	9, 20
<b>Age at amyloid PET, years, Mean (SD)</b>	68.4 (7.5)	65.7 (8.8)
<b>Female, %</b>	73.3	65.5
<b>Education, years, Mean (SD)</b>	16.7 (2.7)	16.5 (2.1)
<b>Cognitive domain impairment, No. 0, 1, 2+</b>	24, 4, 0	17, 4, 1
<b>Vascular risk factors, No. 0-1, 2+</b>	23, 22	19, 10

<b>APOE Genotype</b>	<b><i>APOE3</i></b>	<b><i>APOE4</i></b>
<b>Clinical Dementia Rating (CDR)</b>	0	0
<b>No. of participants</b>	60	37
<b>Age at tau PET, years, Mean (SD)</b>	68.7 (7.9)	64.0 (8.4)
<b>Female, %</b>	66.6	37.8
<b>Education, years, Mean (SD)</b>	16.5 (2.7)	16.6 (2.1)
<b>Cognitive domain impairment, No. 0, 1, 2+</b>	27, 5, 1	15, 6, 1
<b>Vascular risk factors, No. 0-1, 2+</b>	36, 24	23, 14

a, Participants studied for amyloid brain accumulation; b, participants studied for tau accumulation. FBB, participants who received <sup>18</sup>F-Florbetaben; FBP, participants who received <sup>18</sup>F-Florbetapir.

Extended Data Table 3 | *APOE3* and *APOE4* carriers studied for CSF sPDGFR $\beta$  levels

<b>APOE Genotype</b>	<b><i>APOE3</i></b>	<b><i>APOE4</i></b>	<b><i>APOE3</i></b>	<b><i>APOE4</i></b>
<b>Clinical Dementia Rating (CDR)</b>	0	0	0.5	0.5
<b>No. of participants</b>	157	105	40	48
<b>Age at LP, years, Mean (SD)</b>	70.2 (8.9)	67.3 (9.9)	76.5 (7.3)	72.8 (8.1)
<b>Female, %</b>	64	61	39	37
<b>Education, years, Mean (SD)</b>	16.5 (2.7)	16.3 (2.3)	15.7 (2.8)	16.2 (2.8)
<b>Cognitive domain impairment, No. 0, 1, 2+</b>	91, 20, 2	62, 17, 2	9, 8, 14	12, 16, 15
<b>Vascular risk factors, No. 0-1, 2+</b>	100, 57	57, 48	22, 18	18, 30



## Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

### Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- ☐ ☒ The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement
- ☐ ☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- ☐ ☒ The statistical test(s) used AND whether they are one- or two-sided  
*Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
- ☐ ☒ A description of all covariates tested
- ☐ ☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- ☐ ☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- ☐ ☒ For null hypothesis testing, the test statistic (e.g.  $F$ ,  $t$ ,  $r$ ) with confidence intervals, effect sizes, degrees of freedom and  $P$  value noted  
*Give  $P$  values as exact values whenever suitable.*
- ☒ ☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- ☐ ☒ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- ☐ ☒ Estimates of effect sizes (e.g. Cohen's  $d$ , Pearson's  $r$ ), indicating how they were calculated

*Our web collection on [statistics for biologists](#) contains articles on many of the points above.*

### Software and code

Policy information about [availability of computer code](#)

Data collection

The data collection software used were SYNGO (console software) and VE11C (MRI software) for MRI, and PETSyngo 6.7.3 for PET.

Data analysis

Our Rocketship GUI code v1.2 ([https://github.com/petmri/ROCKETSHIP/blob/master/dce/compare\\_gui.m](https://github.com/petmri/ROCKETSHIP/blob/master/dce/compare_gui.m)) running with Matlab R2019a was used for DCE-MRI analyses. We also used FreeSurfer (v5.3.0) software package (<http://surfer.nmr.mgh.harvard.edu/>) for regional brain volume analyses. We used FSL-FLIRT (FMRIB's Linear Image Registration Tool) for PET-MR co-registration (<https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FLIRT>), and Statistical Parametric Mapping (SPM12) was used for PET group comparison in a voxel-by-voxel basis (<https://www.fil.ion.ucl.ac.uk/spm/software/spm12/>). Quantitative Imaging Toolkit was used for data visualization (<https://cabene.io/about/publication/cabene2018quantitative/>). Statistical analyses were conducted with a commercial statistical software package - SPSS (IBM) v26.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

### Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

A full data availability statement is included in the manuscript. The data that support the findings of this study are available from the corresponding author upon reasonable request. See Data Availability statement pasted here: "All data generated and/or analyzed during this study are either included in this article (and its Supplementary Information) or are available from the corresponding author on reasonable request. Source Data for Figs. 1-4 are provided with the article".

# Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

## Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

### Sample size

Based on prior studies demonstrating very large effect sizes for blood-brain barrier markers relating to cognitive impairment (see Montagne et al., Neuron 2015 and Nation et al., Nat Med 2019), we anticipated needing a minimum sample size of 10-20 participants per subgroup (APOE4 status x CDR status) indicating a minimum total sample size of 40-80 to be sufficient to detect significant blood-brain barrier marker effects on cognitive status in all participant subgroups. Consistent with these prior studies, very large effect sizes were observed for all primary study outcomes (up to  $\eta^2 > .32$ ). Based on these observations, a post-hoc power analysis for uncorrected comparisons indicated >99% power to detect significant differences, noncentrality parameter,  $\delta = 5.99$ , critical  $t = 1.99$ , for an  $\alpha = .05$ ). In another post-hoc power analysis for corrected comparisons (ANCOVA with 5 covariates), we again found >99% power to detect significant differences (noncentrality parameter,  $\lambda = 45.57$ , critical  $F = 1.9$ , for an  $\alpha = .05$ ). Post-hoc power analyses were conducted using a general power analysis software program (i.e., G\*Power).

### Data exclusions

Prior to performing statistical analyses, we first screened for outliers using the Grubbs' test, also called the ESD method (Extreme Studentized Deviate), applying a significance level of  $\alpha = 0.01$  (<https://www.graphpad.com/quickcalcs/grubbs1/>). For each of the outliers identified, a secondary index of outlier influence was applied using the degree of deviation from the mean (greater than  $\pm 3$  SDs). Using these stringent criteria, a total of 5 outliers one each in main Figure panels 1j, 1k, and 2j, and one each in Extended Data Figs. 6a and 6b, were removed from analyses, as indicated in the legends of these figures. Continuous variables were also evaluated for departures from normality through quantitative examination of skewness and kurtosis, in addition to visual inspection of frequency distributions. Where departures of normality were identified, log10-transformations were applied, and distribution normalization was confirmed prior to parametric analyses. This has been done for main Fig.s 4h and 4k, and Extended Data Figs. 7a, 7b, 7d, and 7e. Since use of log10-transformations accounts for any non-normality this obviated the need for outliers exclusion.

Further, as stated in our "Participant Inclusion and Exclusion criteria" (see Methods), we included participants ( $\geq 45$  years of age) with neuropsychologically-confirmed no cognitive dysfunction or mild cognitive dysfunction had no current or prior history of any neurological or psychiatric conditions that might better account for any observed cognitive impairment, including organ failure, brain tumors, epilepsy, hydrocephalus, schizophrenia, major depression. Participants were stratified by APOE genotype as APOE4 carriers ( $\epsilon 3/\epsilon 4$  and  $\epsilon 4/\epsilon 4$ ) and APOE4 non-carriers ( $\epsilon 3/\epsilon 3$ ) also defined as APOE3 homozygotes who were cognitively normal or with mild cognitive dysfunction, as determined by clinical dementia rating (CDR) scores and the presence of cognitive impairment in one or more cognitive domains based on comprehensive neuropsychological evaluation including performance on ten neuropsychological tests assessing memory, attention/executive function, language and global cognition. For all analyses individuals with  $\epsilon 3/\epsilon 4$  and  $\epsilon 4/\epsilon 4$  alleles were pooled together in a single APOE4 group, as we did not find in the present cohort consisting of 82-86%  $\epsilon 3/\epsilon 4$  and 14-18%  $\epsilon 4/\epsilon 4$  participants (depending on the outcome measure) a significant difference between the two versus one  $\epsilon 4$  allele for the studied parameters including the BBB permeability unidirectional transfer constant  $K_{trans}$  values and sPDGFR $\beta$  CSF values (see statistical section below). Individuals have been additionally stratified by A $\beta$  and pTau CSF analysis as either A $\beta$ -positive (A $\beta$ 1-42+,  $<190$  pg/mL) or A $\beta$ -negative (A $\beta$ 1-42-,  $>190$  pg/mL), or pTau-positive (pTau+,  $>78$  pg/mL) or pTau-negative (pTau-,  $<78$  pg/mL), using accepted cutoff values.

Participants were excluded if they were diagnosed with vascular cognitive impairment or vascular dementia. Clinical diagnoses were made by neurologists and criteria included whether the patient 1) had a known vascular brain injury and 2) the clinician judged that the vascular brain injury played a role in their cognitive impairment, and/or pattern and course of symptoms. In addition to clinical diagnosis, presence of vascular lesions was confirmed by moderate-to-severe white matter changes and lacunar infarcts by fluid-attenuated inversion recovery (FLAIR) MRI and/or subcortical microbleeds by T2\*-weighted MRI. Participants were also excluded if they were diagnosed with Parkinson's disease, Lewy body dementia or frontotemporal dementia. History of a single stroke or transient ischemic attack was not an exclusion unless it was related to symptomatic onset of cognitive impairment. Participants also did not have current contraindications to MRI and were not currently using medications that might better account for any observed cognitive impairment.

### Replication

Here, we recruited the participants from two sites, University of Southern California (USC), Los Angeles, CA, and Washington University (WashU), St. Louis, MO. At the USC site, participants were recruited through the USC Alzheimer's Disease Research Center (ADRC): combined USC and the Huntington Medical Research Institutes (HMRI), Pasadena, CA. Our previous study findings regarding CSF biomarker analyses were replicated in two independent samples (Nation et al., Nat Med 2019).

The DCE-MRI findings were also conducted at two sites (USC and WashU) using high-resolution 3T MR imaging brain scan protocol. For USC site, a Siemens 3T Prisma scanner was used with a product 32-channel head receive coil and body transmit coil. For WashU site, a Siemens 3T mMR with 20-channel head coil and Siemens 3T Vida with 64-channel head coil were used. The standardization and optimization of the MRI protocol required several tests performed on a phantom. Specifically, scanner characterization and calibration sequences including B0, T1, and variable flip-angle mapping were implemented, optimized, and applied. After the achievement of good results in terms of quality control and reproducibility, we standardized and employed the same pre-contrast and dynamic T1-weighted protocols at both USC and Washington University sites. Of note, all the other MR sequences were identical too on both scanners. In order to minimize the inter-site variability, the entire MRI protocol including the anatomical and DCE pulse sequences were 100% mirrored from one site to another. To minimize inter-site variability, the same contrast agent Gadoterate meglumine (Dotarem®) were injected to participants at the same concentration (0.05 mmol/kg). Finally, the same exact pre- and post-processing analysis pipeline was applied for both sites which includes T1 multi-FA mapping using linear fitting and Patlak-based DCE modeling using the arterial input function determined in each individual from the internal carotid artery. Applying all the above cited factors significantly limited inter-site variability. The consistency of the results from the two sites has been additionally confirmed by our previous publication. In brief, we performed the analysis of the combined DCE datasets from both USC and WashU sites, and additionally site-specific analysis for each of the two sites separately, which showed no statistically significant differences

across sites. Recently, we invited a subset of 52 participants for an additional T1-weighted scan without contrast (using the same scanner and same MR pulse sequences) after their first DCE-MRI and measured both B0 and T1 values at 2-year interval. This study showed that the results were unchanged and consistent across the scans, supporting minimal intra-site variability. USC and WashU DCE-MRI datasets were acquired and processed the same way, and then combined.

The PET scans were also performed at two sites (USC and WashU). At USC site, a Siemens Biograph 64 PET scanner was used. At WashU site, FBP scans were acquired on a Siemens mMR and AV1451 scans were acquired on a Siemens Biograph mCT. The mCT session was used for attenuation correction of the mMR scans. Participants were injected with 300 MBq ( $\pm 10\%$ ) of FBB or 370 MBq ( $\pm 10\%$ ) of FBP. FBB and FBP images were acquired from 90 to 110 min and 50 to 70 min, respectively, after injection in accordance with the manufacturers' recommendation. USC and WashU PET datasets were acquired and processed the same way, and then combined as described in the Methods.

All attempts at replication for CSF, MRI, and PET analyses across sites were successful.

Randomization	We did not randomize since this was not an experimental design.
Blinding	All CSF assays, DCE-MRI and PET scans were conducted by investigators who were blinded to participant demographic information, clinical diagnostic status, and genotype.

## Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

### Materials & experimental systems

n/a	Involved in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input type="checkbox"/>	<input checked="" type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data

### Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input type="checkbox"/>	<input checked="" type="checkbox"/> MRI-based neuroimaging

### Antibodies

Antibodies used	For quantitative western blot assay, the following primary antibody was used: - sPDGFR $\beta$ polyclonal goat IgG antibody (R&D Systems Catalog #AF1042, Lot #GOV0418041). Membranes were incubated with 1 $\mu$ g/mL of antibody overnight at room temperature, then incubated with donkey anti-goat IgG secondary antibody (Invitrogen, Cat #A15999, Lot #44-33-100114, 1:5,000 dilution) for 1 hour at room temperature.
Validation	Recombinant human sPDGFR $\beta$ protein (R&D Systems Catalog #385-PR/CF, Lot #AMO0714072) was used as a positive control in validating the antibodies. Consistently, the manufacture's website specifies this antibody exhibits approximately 35% crossreactivity with recombinant human PDGFR $\beta$ .

### Human research participants

Policy information about [studies involving human research participants](#)

Population characteristics	<p>All participants (n=435) underwent neurological and neuropsychological evaluations performed using the Uniform Data Set and additional neuropsychological tests, and received a venipuncture for collection of blood for biomarker studies. A lumbar puncture was performed in 350 participants (81%) for collection of cerebrospinal fluid (CSF). The dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI) for assessment of blood-brain barrier (BBB) permeability was performed in 245 participants (56%) that had no contraindications for contrast injection. Both LP and DCE-MRI were conducted in 172 participants. Among the 245 DCE-MRI participants, 74 and 96 were additionally studied for brain uptake of amyloid and tau PET radiotracers, respectively.</p> <p>- We studied 245 participants for MRI analysis with clinical dementia ratings of 0 and APOE3 genotype (n=130), 0.5 and APOE3 (n=14), 0 and APOE4 (n=76), and 0.5 and APOE4 (n=25), who exhibited 0 cognitive domain impairment, 1, or 2+ (see Extended Data Table 1 for additional information regarding age, sex, and education).</p> <p>- We studied 74 participants for amyloid PET analysis with clinical dementia ratings of 0 and APOE3 genotype (n=45), and 0 and APOE4 (n=29) (see Extended Data Table 2a for additional information regarding age, sex, and education).</p> <p>- We studied 96 participants for amyloid PET analysis with clinical dementia ratings of 0 and APOE3 genotype (n=60), and 0 and APOE4 (n=37) (see Extended Data Table 2b for additional information regarding age, sex, and education).</p> <p>- We studied 350 participants for CSF sPDGFR<math>\beta</math> analysis with clinical dementia ratings of 0 and APOE3 genotype (n=157), 0.5 and APOE3 (n=40), 0 and APOE4 (n=105), and 0.5 and APOE4 (n=48), who exhibited 0 cognitive domain impairment, 1, or 2+ (see Extended Data Table 3 for additional information regarding age, sex, and education).</p>
Recruitment	<p>Potential participants were identified from the clinics, the community, existing research cohorts or self-referred applications. After an initial prescreening performed by the recruiters, the study coordinators assessed and confirmed the eligibility of the participants before planning and scheduling the procedures and the clinical assessments related to the study.</p>

The multi-modal approach for the identification of potential participants as well as the inclusion of participants from four different sites contributed to minimize the risk of self-selection bias. Furthermore, since the two main tests defining the grouping variables (i.e., the genotyping and the full neuropsychological assessment) are performed only after the enrollment of the participants, the bias related to the recruitment process potentially threatening the validity of the results is negligible. However, it should be noted that higher withdrawal rates were expected in the participants with mild cognitive dysfunction, potentially affecting the statistical power due to an unbalance between the cognitively normal participants and those with mild cognitive impairment.

#### Ethics oversight

University of Southern California (USC) maintains the complete program project and associated project Institutional Review Board (IRB) approvals reviewed and approved by USC IRB (Study IDs HS-16-00759 and HS-15-00875). All other sites also have local IRB approval. Banner University has IRB approval from WIRB (Study number 1173985). Huntington Medical Research Institute has approval from Quorum (QR# 33797/1). Mayo Clinic Arizona has approval from Mayo IRB (IRB#: 1709-97, Study ID PR1709-97-12). Washington University in St. Louis (WashU) has approval from the WashU Human Resources Protection Office (IRB ID#s: 201706100 and 201712150).

Note that full information on the approval of the study protocol must also be provided in the manuscript.

## Magnetic resonance imaging

### Experimental design

Design type	Resting state
Design specifications	N/A
Behavioral performance measures	N/A

### Acquisition

Imaging type(s)	Dynamic Contrast-Enhanced (DCE)-MRI
Field strength	3T
Sequence & imaging parameters	From Methods section: Anatomical coronal spin echo T2-weighted scans were first obtained through the hippocampi (TR/TE 8020/50 ms, NEX = 1, slice thickness 2 mm with 2 mm gap between slices, FOV = 175 x 175 mm, matrix size = 448 x 448). Baseline coronal T1-weighted maps were then acquired using a T1-weighted 3D volumetric interpolated breath-hold sequence (VIBE) sequence and variable flip angle method using flip angles of 2°, 5°, 10°, 12°, and 15°. Coronal DCE-MRI covering the hippocampi and temporal lobes were acquired using a T1-weighted 3D VIBE sequence (FA = 15°, TR/TE = 5.14/2.18 ms, NEX = 1, slice thickness 5 mm with no gap, FOV 175 x 175 mm, matrix size 320 x 320, voxel size was 0.550 x 0.550 x 5 mm <sup>3</sup> ). This sequence was repeated for a total of 16 min with an approximate time resolution of 15.4 sec. Gadolinium-based contrast agent (GBCA), Gadoterate meglumine (Dotarem®, Guerbet, France) (0.05 mmol/kg), was administered intravenously into the antecubital vein using a power injector, at a rate of 3 mL/s followed by a 25 mL saline flush, 30 s into the DCE scan.
Area of acquisition	Scan was coronal with slices covering hippocampi, temporal lobes and other in plane regions until the medial orbital frontal cortex. These regions were chosen due to their importance in aging, cognitive decline, Alzheimer's disease, and prior studies of blood-brain barrier dysfunction in older adults.
Diffusion MRI	<input type="checkbox"/> Used <input checked="" type="checkbox"/> Not used

### Preprocessing

Preprocessing software	<p>From Methods section and Supplementary Information:</p> <ul style="list-style-type: none"> <li>- Blood-brain barrier permeability: Post-processing analysis was performed using Rocketship software running with Matlab. The arterial input function (AIF), which was extracted from a region-of-interest (ROI) positioned at the internal carotid artery (ICA), was fitted with a bi-exponential function prior to fitting with the Patlak model. In a few cases when the ICA was not clearly visible a nearby large arterial vessel was used. The Patlak linearized regression mathematical analysis was used to generate the BBB permeability K<sub>trans</sub> maps, as we previously reported. The high spatiotemporal resolution allowed not only simultaneous measurements of the regional BBB permeability in different white and gray matter regions, but also accurate calculations of the K<sub>trans</sub> values in small anatomical regions as thin as cortical gray matter areas. The present analysis requires that the tracer's diffusion across the BBB remains unidirectional during the acquisition time. The total tracer concentration in the tissue, C<sub>tissue</sub> (t), can be described as a function of the blood concentration, C<sub>AIF</sub> (t), the intravascular blood volume, v<sub>p</sub>, and a blood-to-brain transfer constant, K<sub>trans</sub>, that represents the flow from the intravascular to the extravascular extracellular space using equation (see Methods). A statistically significant intersubject variability in the measurement of v<sub>p</sub> was not observed. ROI-averaged analysis of DCE-MRI output maps was performed by an experienced neuroradiologist who manually drew ROIs for each participant based on their own anatomy since a substantial variability between individuals is seen at a macroscopic level (e.g., enlarged ventricles, cortical atrophy, hippocampal shrinkage, etc.). Thus, the regional BBB K<sub>trans</sub> permeability were measured in 10 different gray matter ROIs including the hippocampus (HC), parahippocampal gyrus (PHG), caudate nucleus, thalamus, striatum, orbital frontal cortex (OFC), and inferior temporal gyrus (ITG), and white matter ROIs including subcortical watershed white matter fibers, corpus callosum, and internal capsule.</li> <li>- Volumetric analysis: HC and PHG morphometry were performed using the FreeSurfer (v5.3.0) software package (<a href="http://surfer.nmr.mgh.harvard.edu/">http://surfer.nmr.mgh.harvard.edu/</a>), as previously performed. HC and PHG were segmented using FreeSurfer Desikan-</li> </ul>
------------------------	--



Killiany and subcortical atlases. Then, regional volumes (mm<sup>3</sup>) were derived accordingly. The technical details of this procedure are described in previous publications. Data processing and visualization were performed using the Laboratory of Neuro Imaging (LONI) pipeline system (<http://pipeline.loni.usc.edu>) and Quantitative Imaging Toolkit.

Normalization

N/A

Normalization template

N/A

Noise and artifact removal

Motion correction (for DCE-MRI) was applied using ImageJ's Stack Reg - Rigid Body plugin.

Volume censoring

N/A

## Statistical modeling & inference

Model type and settings

N/A

Effect(s) tested

N/A

Specify type of analysis: ☐ Whole brain ☒ ROI-based ☐ Both

Anatomical location(s)

ROI-averaged analysis of DCE-MRI output maps was performed by an experienced neuroradiologist who manually drew ROIs for each participant based on their own anatomy since a substantial variability between individuals is seen at a macroscopic level (e.g., enlarged ventricles, cortical atrophy, hippocampal shrinkage, etc.). Thus, the regional BBB Ktrans permeability were measured in 10 different gray matter ROIs including the hippocampus (HC), parahippocampal gyrus (PHG), caudate nucleus, thalamus, striatum, orbital frontal cortex (OFC), and inferior temporal gyrus (ITG), and white matter ROIs including subcortical watershed white matter fibers, corpus callosum, and internal capsule. Regions were chosen to cover a variety of brain regions and tissue compartments (e.g., cortical white matter, subcortical white matter, cortical grey matter, subcortical grey matter, and limbic areas of special focus--in this case the hippocampus and medial temporal lobes). The regional brain volumes were measured in HC and PHC gyri.

Statistic type for inference  
(See [Eklund et al. 2016](#))

We did not conduct a voxel-level analysis.

Correction

We did not conduct a voxel-level analysis.

## Models & analysis

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Functional and/or effective connectivity
<input checked="" type="checkbox"/>	<input type="checkbox"/> Graph analysis
<input checked="" type="checkbox"/>	<input type="checkbox"/> Multivariate modeling or predictive analysis

# Injured adult neurons regress to an embryonic transcriptional growth state



<https://doi.org/10.1038/s41586-020-2200-5>

Received: 12 April 2019

Accepted: 13 February 2020

Published online: 15 April 2020

 Check for updates

Gunnar H. D. Poplawski<sup>1</sup>, Riki Kawaguchi<sup>2,3</sup>, Erna Van Niekerk<sup>1</sup>, Paul Lu<sup>1,4</sup>, Neil Mehta<sup>1</sup>, Philip Canete<sup>1</sup>, Richard Lie<sup>1</sup>, Ioannis Dragatsis<sup>5</sup>, Jessica M. Meves<sup>1</sup>, Binhai Zheng<sup>1,4</sup>, Giovanni Coppola<sup>2,3</sup> & Mark H. Tuszynski<sup>1,4</sup>

Grafts of spinal-cord-derived neural progenitor cells (NPCs) enable the robust regeneration of corticospinal axons and restore forelimb function after spinal cord injury<sup>1</sup>; however, the molecular mechanisms that underlie this regeneration are unknown. Here we perform translational profiling specifically of corticospinal tract (CST) motor neurons in mice, to identify their ‘regenerative transcriptome’ after spinal cord injury and NPC grafting. Notably, both injury alone and injury combined with NPC grafts elicit virtually identical early transcriptomic responses in host CST neurons. However, in mice with injury alone this regenerative transcriptome is downregulated after two weeks, whereas in NPC-grafted mice this transcriptome is sustained. The regenerative transcriptome represents a reversion to an embryonic transcriptional state of the CST neuron. The huntingtin gene (*Htt*) is a central hub in the regeneration transcriptome; deletion of *Htt* significantly attenuates regeneration, which shows that *Htt* has a key role in neural plasticity after injury.

The corticospinal tract (CST), which arises from the cerebral cortex and projects to the spinal cord, is the most important projection for voluntary movement in humans<sup>2,3</sup>. The failure of the CST to regenerate has been a major limiting factor in the advancement of potential regenerative therapies for the treatment of spinal cord injuries in humans<sup>3,4</sup>. Two experimental manipulations substantially enhance the growth of injured corticospinal neurons: (1) deletion of the tumour-suppressor genes *Pten* and *Socs3*<sup>5,6</sup>, and (2) grafts of homologous neural stem cells or NPCs into sites of spinal cord injury<sup>1,7</sup>. Deletion of *Pten* alone, or of *Pten* and *Socs3*, enables regeneration and/or sprouting only through small strands of host glia<sup>5,6</sup>, whereas NPC grafts enable robust regeneration of corticospinal axons into the lesion site itself<sup>1</sup>. Untreated injured corticospinal axons without a growth-promoting NPC graft completely lack regeneration into the lesion site<sup>1</sup> (Extended Data Fig. 10a). To date, the intrinsic neuronal mechanisms that are recruited by neural stem cell grafts and enable robust corticospinal regeneration in the host have not been identified.

Here we identify transcriptional mechanisms that are associated with corticospinal regeneration. We took advantage of the ability of NPC grafts to support corticospinal regeneration, together with a genetic mouse model (Glt25d2-eGFP-L10a mice, which express enhanced green fluorescent protein (eGFP)-tagged polyribosomes in layer-5b cortical neurons) to selectively enrich actively translated mRNA pools from layer-5b cortical projection neurons (including corticospinal neurons)<sup>8</sup>. We report the regenerative transcriptome of the corticospinal neuron. To our surprise, this regeneration state is triggered by injury alone, driving the corticospinal neuron to a transcriptional profile that is immature; NPC grafts successfully sustain the corticospinal regenerative transcriptome over time, rather than recruiting a new growth

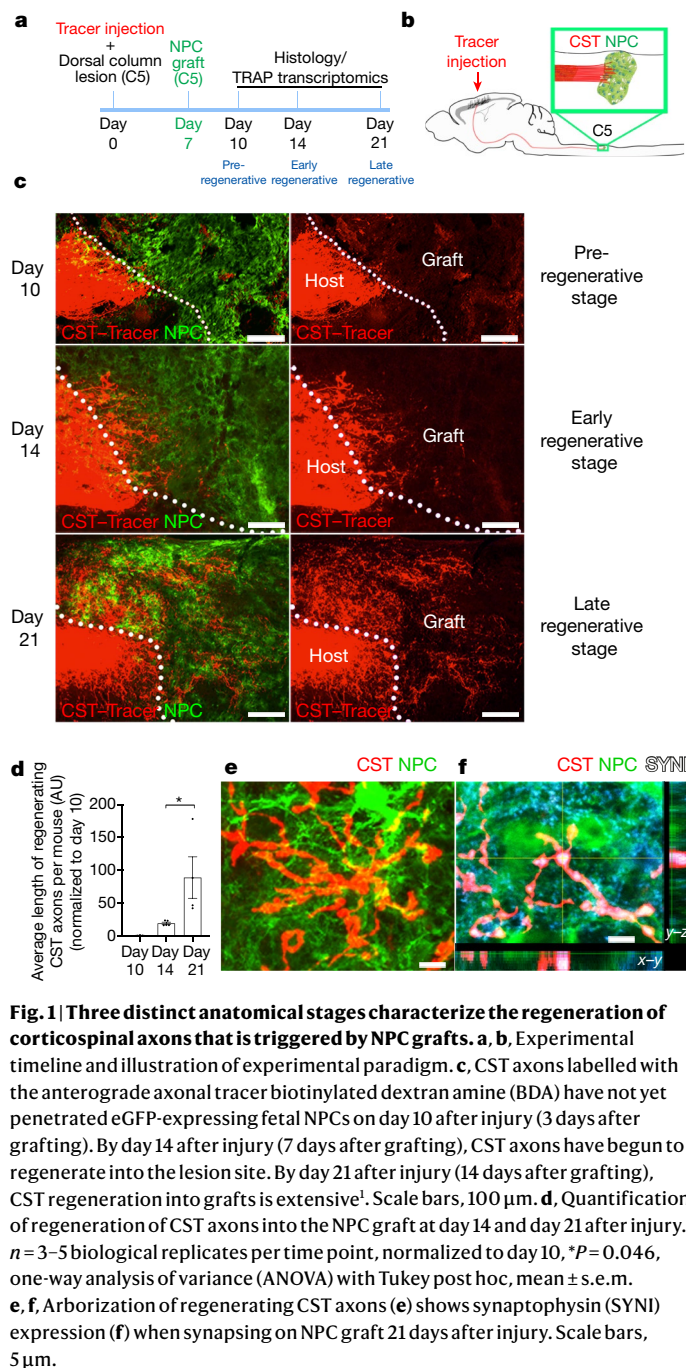
program in neurons. We identify *Htt* as an essential regulatory hub in the corticospinal regenerative state, and *Htt*-knockout mice exhibit substantially reduced corticospinal sprouting and regeneration.

## Three stages of CST regeneration

It has previously been reported that grafts of spinal-cord-derived NPCs, or embryonic stem cells driven to neural stem cells of spinal cord identity, support robust regeneration of corticospinal axons after spinal cord injury<sup>1,9,10</sup>. The availability of the transgenic Glt25d2-eGFP-L10a mouse line<sup>8</sup> provided a tool for identifying the transcriptomic mechanisms that are associated with corticospinal regeneration.

We performed dorsal column lesions at cervical spinal cord level 5 (C5) to transect the corticospinal projection in adult Glt25d2-eGFP-L10a mice<sup>8</sup>. One week later, we grafted eGFP-expressing NPCs derived from the embryonic-day-12 (E12) mouse spinal cord into the lesion (Fig. 1a, b). Analysis of anterogradely labelled corticospinal axons at 10, 14 and 21 days after injury (3, 7 and 14 days after grafting) illustrated three anatomically distinct stages of the regeneration of corticospinal axons into grafts (Fig. 1c, d). (1) In the ‘pre-regenerative’ stage at 10 days after the lesion, corticospinal axons had not yet penetrated NPC grafts. The axons were still retracted within 1 mm of the lesion site. (2) Four days later, we observed an ‘early regenerative’ stage in which CST axons penetrated the proximal region of the NPC graft. At this time, axons extended approximately 100 µm into grafts but lacked dense arborization. (3) In the ‘late regenerative’ stage at 21 days after the lesion, corticospinal axons extensively regenerated 0.5–1 mm into NPC grafts, arborized extensively and formed synapses onto grafted neurons (Fig. 1e, f). We accordingly characterized the transcriptional state at these three time

<sup>1</sup>Department of Neurosciences, University of California San Diego, La Jolla, CA, USA. <sup>2</sup>Department of Psychiatry, University of California Los Angeles, Los Angeles, CA, USA. <sup>3</sup>Department of Neurology, University of California Los Angeles, Los Angeles, CA, USA. <sup>4</sup>Veterans Administration Medical Center, San Diego, CA, USA. <sup>5</sup>Department of Physiology, University of Tennessee, Memphis, TN, USA. ✉e-mail: [gunnar.poplawski@gmail.com](mailto:gunnar.poplawski@gmail.com); [mtuszynski@ucsd.edu](mailto:mtuszynski@ucsd.edu)



**Fig. 1 | Three distinct anatomical stages characterize the regeneration of corticospinal axons that is triggered by NPC grafts.** **a, b,** Experimental timeline and illustration of experimental paradigm. **c,** CST axons labelled with the anterograde axonal tracer biotinylated dextran amine (BDA) have not yet penetrated eGFP-expressing fetal NPCs on day 10 after injury (3 days after grafting). By day 14 after injury (7 days after grafting), CST axons have begun to regenerate into the lesion site. By day 21 after injury (14 days after grafting), CST regeneration into grafts is extensive<sup>1</sup>. Scale bars, 100  $\mu$ m. **d,** Quantification of regeneration of CST axons into the NPC graft at day 14 and day 21 after injury.  $n = 3-5$  biological replicates per time point, normalized to day 10,  $^*P = 0.046$ , one-way analysis of variance (ANOVA) with Tukey post hoc, mean  $\pm$  s.e.m. **e, f,** Arborization of regenerating CST axons (**e**) shows synaptophysin (SYN) expression (**f**) when synapsing on NPC graft 21 days after injury. Scale bars, 5  $\mu$ m.

points (10, 14 and 21 days after injury), representing states of pre-, early and late regeneration.

## TRAP of regenerating CST neurons

We characterized the transcriptional profile of regenerating versus non-regenerating adult corticospinal neurons in Glt25d2-eGFP-L10a mice by isolating the cellular pool of actively translated mRNAs using translating ribosome affinity purification (TRAP)<sup>11</sup>. In intact Glt25d2-eGFP-L10a mice, eGFP was expressed specifically in layer-5b cortical pyramidal neurons (Extended Data Fig. 1a, b), including corticospinal neurons (as previously reported<sup>8,12</sup>). Thus, dissection of the motor cortex in these mice enables the identification of mRNAs that are highly enriched in corticospinal motor neurons: indeed, approximately 50% at least of layer-5b pyramidal neurons in these mice are corticospinal, identified by double labelling for eGFP and the retrograde

neuronal tracer cholera toxin B injected into the spinal cord at cervical levels (Extended Data Fig. 1c, d). mRNA that was affinity-purified from the motor cortex was therefore enriched in corticospinal neurons that project into the cervical spinal cord. We performed TRAP and RNA sequencing (RNA-seq) of the motor cortex at 10, 14 and 21 days after injury (Fig. 1a, b, Extended Data Fig. 1e)

## RNA-seq of regenerating CST neurons

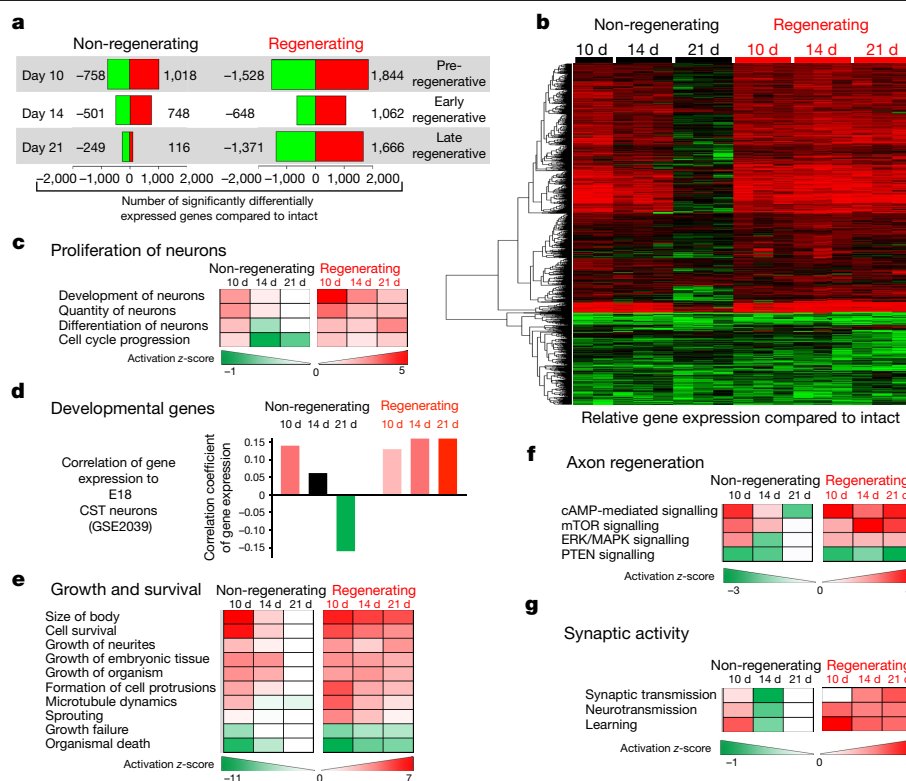
Gene-expression correlation analysis (Extended Data Fig. 2) showed that mRNA fractions isolated from eGFP-tagged polyribosomes (hereafter referred to as 'TRAPed' samples) were distinct in their transcriptomic profiles from sample mRNA fractions that were not bound to eGFP-affinity-purified samples (hereafter, unbound samples), which confirmed the successful purification of TRAPed mRNA from layer-5b cortical neurons. The TRAPed fraction of intact mice was several-fold enriched in genes specifically expressed in neurons of cortical layer 5b (as annotated by the Allen Mouse Brain Atlas<sup>13</sup>), including *Fezf2*<sup>14</sup>, *Parm1*<sup>14</sup> and *Crym*<sup>15,16</sup> (Extended Data Fig. 3), and was depleted of glial genes as expected from an enriched neuronal sample (Extended Data Fig. 4). To further investigate the validity of the RNA-seq dataset, we confirmed that three sample genes that were differentially regulated upon regeneration of corticospinal axons (*Sox9*, *Pax2* and *Atg7*) exhibited corresponding changes in protein expression (Extended Data Fig. 5).

## Regeneration prolongs the injury transcriptome

Corticospinal regeneration is associated with the persistence of the neuronal injury signal: compared to intact mice, a total of 3,372 and 1,776 genes were significantly differentially expressed in datasets from regenerating and lesioned control (non-regenerating) mice at the pre-regeneration time point (10 days after injury), respectively, using a false discovery rate (FDR) of  $P < 0.1$ . There were 1,710 and 1,249 genes that were differentially expressed at 14 days after injury. At the late regeneration time point (21 days after injury), 3,037 and 365 genes were differentially expressed in the regenerating and lesioned control datasets, respectively (Fig. 2a). Thus, regenerating corticospinal neurons maintained an active transcriptomic program in the presence of NPC grafts over the entire sampled time period (3,372, 1,710 and 3,037 differentially regulated genes at time points of 10, 14 and 21 days after injury, respectively), whereas the early rise in differentially expressed genes was progressively attenuated in lesioned, non-regenerating neurons (1,776, 1,249 and 365 genes at 10, 14 and 21 days after injury, respectively).

A notable pattern that emerged from inspection of heat maps of these data was that many of the most-highly differentially expressed genes were altered in parallel in lesioned control and regenerating animals at the 10-day, pre-regenerative time point (Fig. 2b, Extended Data Figs. 6, 7). Many of these differentially expressed genes (compared to intact subjects) represented stem-cell-like gene families by Gene Ontology analysis (Fig. 3c), as well as the activation of proliferation-, differentiation-, and cell-cycle-progression-related functions (Fig. 2c, Extended Data Fig. 7). These include the upregulation of genes that have previously been shown to induce direct neuronal differentiation, such as *Sox2* and *Ascl1*<sup>17</sup>. This finding suggests that one of the consequences of injury is the reversion or transcriptional reprogramming of the adult injured corticospinal neuron to a more-immature state.

At 14 days after injury, when regeneration into grafts was beginning, many of the same genes that were upregulated in response to injury remained expressed in regenerating mice, but were beginning to lose expression in lesioned controls (Fig. 2b, Extended Data Fig. 6). Notably, one week later (21 days after injury)—when axonal regeneration was robust in NPC-grafted animals and failed in lesioned controls—only regenerating mice maintained expression of many of the same sets of



**Fig. 2 | Transcriptomic analysis of regenerating corticospinal neurons reveals injury-induced reversal to an immature developmental state that is sustained upon regeneration.** **a**, Number of significantly differentially expressed genes (DEGs) of injured (non-regenerating) and regenerating CST neurons relative to intact CST neurons (FDR < 0.1, Benjamini–Hochberg, two-sided) at days 10, 14 and 21 after injury. **b**, Heat map of top 1,000 DEGs (FDR < 0.1, Benjamini–Hochberg, two-sided) upon injury and upon regeneration relative to intact neurons, arranged by hierarchical clustering. Red, increased expression; green, reduced expression. The intensity of the colour reflects the degree of gene regulation. **c–g**, Pathway analysis of DEGs from **b**. The state of activity of a signalling pathway or biological function is represented by a z-score (right-tailed Fisher’s exact test). Red, activation; green, inhibition. Individual scaling parameters are listed at the bottom of each graph. **c**, Sustained activation of proliferation and differentiation pathways, only in regenerating CST neurons. **d**, Correlation analysis of DEGs between our

genes that were activated as a function of initial injury (Fig. 2b, Extended Data Figs. 6, 7).

To account for potential changes in gene expression due to physical contact of axons with exogenous cell grafts (in the absence of axonal regeneration), an additional control group received grafts of bone-marrow stromal cells, which are non-permissive to CST regeneration<sup>1</sup>. At 21 days after injury, gene expression in the bone-marrow stromal cell group was comparable to the lesion-only group without cell grafting (Extended Data Fig. 8a, b). These findings further demonstrate that the successful regeneration of the corticospinal system is accompanied by sustained activation of transcriptional programs (Figs. 2, 3).

## Reversion to a neural development state

Corticospinal regeneration is associated with the reversion to a developmental state of gene expression: we compared the transcriptomic profiles of regenerating CST neurons to a previously published developmental dataset of corticospinal neurons<sup>18</sup> (Methods). We found a significant correlation ( $c = 0.16$ ,  $P < 6 \times 10^{-29}$ ) between the transcriptomic profiles of regenerating corticospinal neurons from our regenerating dataset and E18 corticospinal neurons (Fig. 2d). This positive correlation was sustained throughout all time points in the regeneration group,

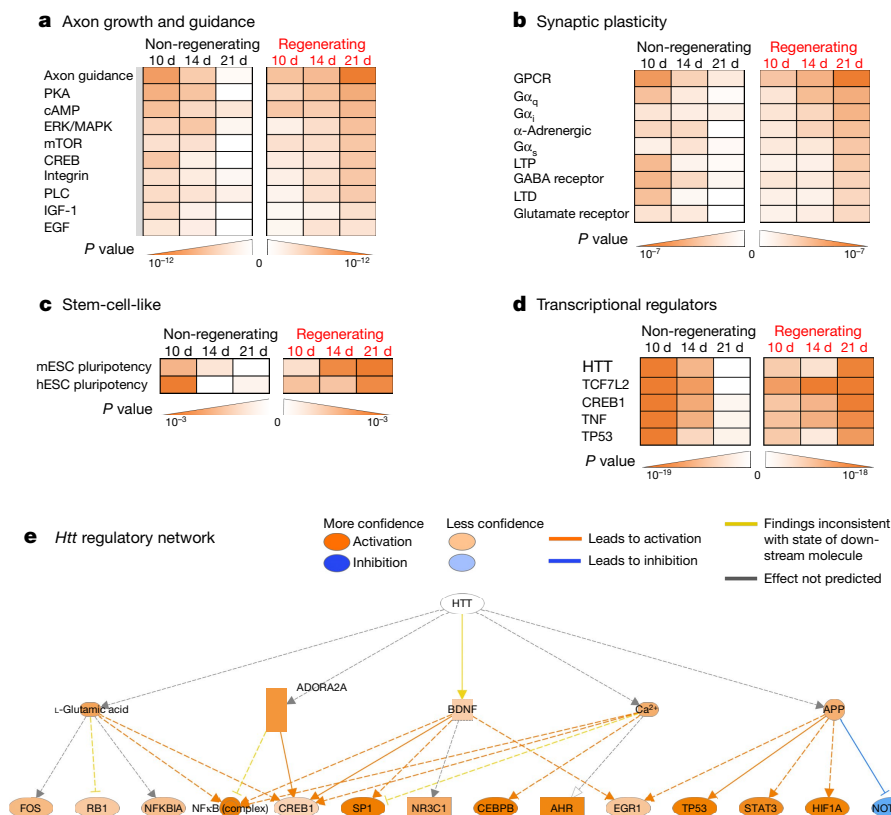
dataset and embryonic (E18) CST neurons from a previous study (Gene Expression Omnibus (GEO) accession code GSE2039). Gene expression was ranked by log-transformed fold change for each dataset individually, and then correlated via rank–rank hypergeometric overlap. Correlation coefficients for each condition compared to the embryonic CST dataset are shown. Regenerating CST neurons correlate positively (red) with gene expression of embryonic CST neurons, whereas in the lesion-only group at 21 days we observed negative correlation (green). **e**, Sustained activation of cellular growth and sustained inhibition of apoptotic pathways, only in regenerating CST neurons. **f**, Increased axon regeneration is reflected in the activation of signalling pathways that promote axon regeneration (for example, mTOR) as well as inhibition of growth-inhibitory signalling pathways (PTEN). **g**, Synaptic activity and rewiring of circuits is activated through all time points of CST regeneration.  $n = 2$  biological replicates for day-10 non-regenerating and  $n = 3$  biological replicates for all other conditions.

but in lesion-only controls shifted from a positive correlation at 10 days after injury ( $c = 0.14$ ,  $P < 8 \times 10^{-22}$ ) to a negative correlation at 21 days after injury ( $c = -0.16$ ,  $P < 2 \times 10^{-26}$ ) (Fig. 2d). Moreover, compared to the strong correlation of our regenerating dataset to E18 corticospinal neurons, the correlation progressively diminished when compared to more-mature corticospinal neurons at postnatal day 3 ( $c = 0.11$ ,  $P < 9 \times 10^{-13}$ ) and postnatal day 6 ( $c = 0.10$ ,  $P < 2 \times 10^{-11}$ ) (Extended Data Fig. 8c), which suggests that the state of regeneration most closely correlates with an embryonic transcriptional state. This supports the hypothesis that injured corticospinal neurons initially revert to an immature, embryonic transcriptomic state, which is sustained over prolonged time points as regeneration proceeds.

## Unique transcriptional networks over time

Gene Ontology analysis indicated that corticospinal regeneration was associated with an activated state of cellular growth and suppression of growth-inhibitory and apoptotic pathways (Fig. 2e). Significantly enriched signalling pathways involved in regeneration relate to classic axon growth and guidance molecules, and synaptic plasticity (Fig. 2f, g). Several genes (including members of the Erk family<sup>19</sup>, *Mtor*<sup>5,20</sup> and *Igf1*<sup>21</sup>) that have previously been identified as linked to regeneration were





**Fig. 3 | Transcriptomic analysis of regenerating corticospinal neurons identifies *Htt* as a potential upstream regulator of regeneration.**

**a–d**, Pathway analysis was performed using all DEGs relative to gene expression in intact mice (FDR < 0.1). The significance of a signalling pathway or biological function represented by a *P* value using a right-tailed Fisher's exact test is shown; darker orange denotes a higher significance. Individual scaling parameters are listed at the bottom of each graph. The *P* value (right-tailed Fisher's exact test) represents the significance of overlap between DEGs within each group and with genes classified as being involved in a particular category. **a**, The most significantly involved pathway in this dataset is axon growth and guidance, with increasing involvement over time in the regenerating cohort and complete decline in the non-regenerating cohort. **b**, Genes involved in

synaptic plasticity become more significantly involved at 21 days after lesion in the regenerating cohort. **c**, DEGs involved in the pluripotency of stem cells are more significantly enriched over time as CST neurons regenerate, and less enriched as regeneration fails. hESC, human embryonic stem cell; mESC, mouse embryonic stem cell. **d**, Prediction of transcriptional regulators identifies *Htt* as an important potential upstream regulator of CST regeneration ( $P < 10 \times 10^{-17}$ , right-tailed Fisher's exact test for 21-day regenerating CST). **e**, The *Htt* regulatory network places *Htt* as a regulatory hub in the regulation of genes involved in axon regeneration (for example, *Bdnf*, *Trp53* and *Stat3*).  $n = 2$  biological replicates for day 10 non-regenerating and  $n = 3$  biological replicates for all other conditions.

included among the genes that we found to be associated with regeneration (Fig. 3a), supporting the validity of this dataset. We also identified several additional genes that are potentially related to regeneration, as summarized below.

The gene-expression profiles in the pre-regenerative (day-10) state that were distinct from those of intact mice included cell survival, proliferation (Fig. 2c, e), metabolism and axon guidance (Extended Data Fig. 8d), amongst which were Wnt genes, *Ctnnb1* and genes involved in cAMP and PKA signalling. Genes in these datasets that were significantly differentially regulated included *L1cam*, *Wnt5a*, *Wnt7b* and the netrin receptor *Unc5b*. We also observed upregulation of growth factors (including *Igf1* and *Tgfa*) and of transcription factors related to neuronal differentiation and axon regeneration (including *Sox2*, *Ascl1*, *Klf3*, *Klf7*, *Satb1* and *Trp53* (also known as *Trp53*)) as well as downregulation of *Junb*, *Klf2* and *Atf2* (Supplementary Table 1). As noted above, many of these genes exhibited continued expression over the 21-day period during which regeneration was successful. Other genes that were significantly and persistently expressed with regeneration included proteins involved in axon guidance, such as semaphorins (*Sema3a* and *Sema5b*); genes involved in axonal transport (such as *Kif5a*, *Uchl1* and *Dlg2*); and genes for cytoskeletal proteins (such as *Tubb2b*); as well as the upregulation of *Nefl* and downregulation of *Nefh* (Supplementary Table 1).

In the early regenerative (day-14) state, transcriptional networks related to protein translation, energy production and mitochondrial function were activated (Extended Data Fig. 8e), suggesting increased cellular activity to support axonal growth. We also observed the activation of genes involved in classic regenerative pathways, such as *Mtor*<sup>5,20</sup> (Fig. 2f, Extended Data Fig. 8e), and genes involved in PKA signalling<sup>22</sup> (Fig. 3a, Extended Data Fig. 8e). The activation of molecules known to be involved in the neuronal growth state and axon pathfinding also included *Npy*<sup>23</sup>, *Rock2*<sup>24</sup> and *Plxnb3*<sup>25</sup> (Supplementary Table 2). Specific individual genes differentially expressed at this time point included *Nrp2*, *Map2k1* and *Ptprz1* (Supplementary Table 2)—molecules that have previously been identified as important in central axonal regeneration<sup>26–29</sup>.

By the late regenerative state (day 21), we observed the upregulation of integrin genes such as *Itga4*, which could be involved in axon pathfinding; *Stx12*; genes for kinases, including *Cdk1*, *Pak7* and *Mapk3*; and axon guidance molecules, including upregulation of the genes *Ntn1* and *Ntf3*; and downregulation of semaphorin genes, including *Sema3c*, *Sema3e* and *Sema7a* (Supplementary Table 3). Specific genes differentially expressed at this time point included *Klf6*, *Igf2*, the GABA<sub>A</sub> receptor genes, *Kif5b* and *Myo5b* (Supplementary Table 3). Voltage-gated ion channels, including *Kcna4* and *Kcna5*, were also upregulated in the late regenerative state, possibly reflecting the establishment of functional

synapses (Supplementary Table 3). The late regenerative phase corresponds to a stage in which regenerating axons are actively guided by and synapsing on neurons within the NPC grafts<sup>1,30–32</sup>, and have shifted to a synaptogenic phase of growth (Figs. 2g, 3b, Extended Data Fig. 8f).

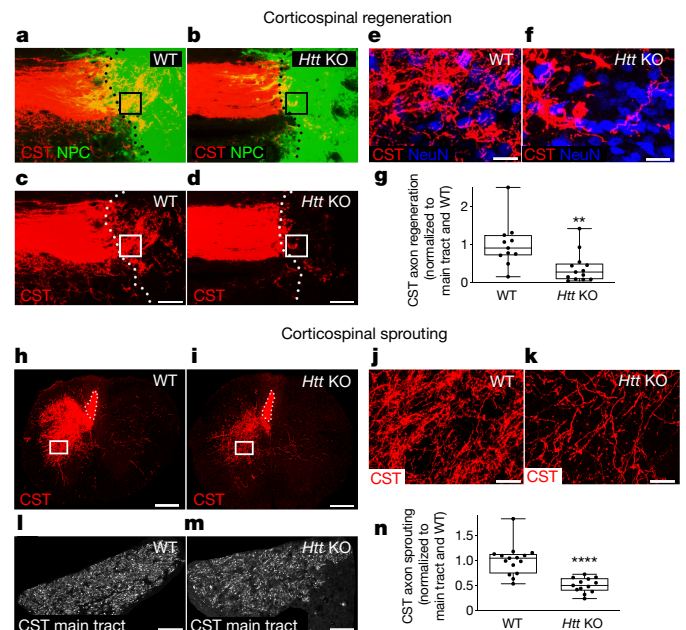
Thus, successful corticospinal regeneration is characterized by at least two aspects of the transcriptional profile: (1) a reversion to an immature state, which persists as regeneration is underway, and (2) temporally shifting patterns of gene expression, in which anti-apoptotic cellular mechanisms are activated early after injury, shifting to the expression of genes that contribute to an active early regenerative state (including *Pten* suppression and *Mtor* activation<sup>33</sup> (Fig. 2f)) and finally—as axons have extended deeply into grafts and are forming synapses<sup>1,9,31</sup>—the expression of axon guidance molecules and synaptic genes (Figs. 2g, 3b).

### *Htt* is essential for CST regeneration

Transcriptional profiling can reveal previously unidentified mechanisms that mediate biological phenomena, such as axonal regeneration. *Htt* arose in our analysis as an unexpected and potentially key regulatory hub in axonal regeneration (Fig. 3d), among the top three networks that were statistically associated with regeneration. Members of the *Htt* regulatory network were enriched six-fold in cortical samples associated with early axonal injury ( $P < 0.001$ ) compared to late injury, and persisted at significantly elevated levels only in regenerating conditions ( $P < 0.001$ ) (Fig. 3d, Extended Data Fig. 8e, Supplementary Table 4). Moreover, *Htt* occupied a central role as a hub in the regenerating transcriptome (Fig. 3e), potentially influencing regeneration genes including *Nfkb* (also known as *Nfkb1*), *Stat3* and *Bdnf* (Fig. 3e). Furthermore, *Htt* has potential functions in the development of cortical projections<sup>34,35</sup>, consistent with transcriptomic analyses that suggest that the injured or regenerating neuron reverts to an immature state. Although expansions of CAG trinucleotide repeats in *Htt* are associated with the development of Huntington's disease, the typical role of *Htt* is incompletely understood. *Htt*-knockout mice exhibit early embryonic lethality<sup>36</sup>, whereas the effects of conditional *Htt* deletion are somewhat controversial. One study has shown motor coordination abnormalities one month after conditional *Htt* deletion in adult mice<sup>37</sup>, whereas another study found no behavioural abnormalities up to seven months after conditional *Htt* deletion<sup>38</sup>. *Htt* is a regulator of *Creb1*<sup>39</sup> and *Tp53*<sup>40</sup> (Fig. 3d, e), and has previously been implicated in neurogenesis<sup>38,41</sup>, production and transport of BDNF<sup>42</sup>, NFkB expression<sup>43</sup> and axonal transport<sup>44</sup>. Given the identification of *Htt* as a central transcriptional mediator of axonal regeneration, we performed additional studies using mice with conditional *Htt* deletion.

Adeno-associated virus (AAV)8 vectors expressing tdTomato and Cre recombinase were injected into the right motor cortices of four-month-old *Htt<sup>fllox/fllox</sup>* mice<sup>45</sup> (Fig. 4), a time point that could not affect neuronal development<sup>38</sup>. Six weeks later, we performed C4 spinal-cord dorsal-column lesions, and grafted eGFP-expressing E12 spinal-cord-derived NPCs into the lesion. A separate group of *Htt<sup>fllox/fllox</sup>* mice received AAV injections that lacked *cre*, followed by NPC grafts. Three mice that were euthanized six weeks after injection of AAV8-tdTomato-Cre confirmed the elimination of detectable HTT protein (Extended Data Fig. 9). Four weeks after graft placement, corticospinal regeneration into grafts was reduced by 60% in mice with *Htt* deletion compared to *Htt<sup>fllox/fllox</sup>* mice that lacked Cre injections ( $P < 0.01$ , two-tailed *t*-test) (Fig. 4a–g). Thus, *Htt* is necessary for corticospinal regeneration, which identifies a role for *Htt* in neuronal repair.

In the same mice, we also determined whether *Htt* deletion influenced spontaneous corticospinal sprouting into host grey matter above the lesion<sup>46–48</sup>. Indeed, *Htt* deletion resulted in a 50% reduction in sprouting at C2 after C4 spinal-cord lesions, compared to *Htt<sup>fllox/fllox</sup>* mice that lacked Cre injections<sup>49</sup> ( $P < 0.0001$ , two-tailed *t*-test) (Fig. 4h–n, Extended Data Fig. 10b).



**Fig. 4 | *Htt* is essential for regeneration of corticospinal axons.** *Htt* was conditionally deleted in the cortical spinal projections of the right motor cortex. **a–d**, CST axon regeneration is reduced in NPC grafts upon host *Htt* deletion. WT, wild type. Scale bars, 500  $\mu$ m. **e, f**, NeuN-positive graft areas are less penetrated by *Htt*-deficient CST axons (**f**) compared to *Htt*-positive CST axons (**e**). Scale bars, 20  $\mu$ m. **g**, Quantification of total axon length in pixel in graft area (normalized to individual axon counts in the main tract, and then normalized to axon regeneration in the wild-type condition) shows a significant reduction in the regeneration of CST axons upon host *Htt* deletion. \*\* $P = 0.0064$ , unpaired two-tailed *t*-test ( $n = 11$  wild type;  $n = 13$  *Htt* knockout). All quantifications were normalized to total axon counts in the main CST in each subject. **h–m**, We also assessed sprouting of the CST into host cervical grey matter rostral to the lesion site at C2. The boxed areas in **h** and **i** are shown magnified in **j** and **k**, respectively; and the areas with dotted outlines in **h** and **i** are shown magnified in **l** and **m**, respectively. Scale bars, 200  $\mu$ m (**h, i**); 25  $\mu$ m (**j, k**); 30  $\mu$ m (**l, m**). **n**, *Htt* deletion in the host significantly reduced sprouting of CST axons. \*\*\*\* $P < 0.0001$ , unpaired two-tailed *t*-test, ( $n = 15$  wild type;  $n = 13$  *Htt* knockout). All quantifications were normalized to total axon counts in the main CST in each subject. In the box plots, centre line is the median, box extends from the 25th to the 75th percentile and whiskers denote minimum–maximum values.

### Discussion

The regenerating corticospinal motor neuron transcriptome represents a fundamental shift in state towards an immature cell, and requires *Htt* to support successful repair. Notably, the transcriptional response to injury is nearly identical in the presence or absence of a regeneration-enabling cell graft, yet the pro-growth, immature cell state persists only when conditions permissive to growth (for example, a stem cell implant in the lesion site) are present.

The transcriptomic signature associated with regeneration evolves over time points that sample pre-, early and late regeneration states (Figs. 2, 3, Extended Data Fig. 8d–f). In pre-regeneration, many genes are pro-survival; subsequently, expressed genes have protein products that are associated with the axonal growth state (for example, mTOR) and axonal elongation (for example, cAMP, ERK and MAPK); finally, late regeneration genes reflect axon guidance and synaptogenesis. Thus, in a two-week time frame, classical neurodevelopmental events occur that are associated with successful regeneration.

Many transcriptional regulators related to corticospinal regeneration were identified in this study, including *Htt*. Indeed, conditional gene deletion showed that *Htt* is required for neuronal repair. Throughout life, neuronal maintenance and repair are essential to support

adequate cellular functioning; thus, *Htt* mutations are predicted to impair neuronal repair mechanisms—eventually resulting in neuronal degeneration. This is a testable hypothesis for future studies.

## Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-020-2200-5>.

- Kadoya, K. et al. Spinal cord reconstitution with homologous neural grafts enables robust corticospinal regeneration. *Nat. Med.* **22**, 479–487 (2016).
- Nathan, P. W. Effects on movement of surgical incisions into the human spinal cord. *Brain* **117**, 337–346 (1994).
- Friedli, L. et al. Pronounced species divergence in corticospinal tract reorganization and functional recovery after lateralized spinal cord injury favors primates. *Sci. Transl. Med.* **7**, 302ra134 (2015).
- Oudega, M. & Perez, M. A. Corticospinal reorganization after spinal cord injury. *J. Physiol. (Lond.)* **590**, 3647–3663 (2012).
- Liu, K. et al. PTEN deletion enhances the regenerative ability of adult corticospinal neurons. *Nat. Neurosci.* **13**, 1075–1081 (2010).
- Jin, D. et al. Restoration of skilled locomotion by sprouting corticospinal axons induced by co-deletion of PTEN and SOCS3. *Nat. Commun.* **6**, 8074 (2015).
- Coumans, J. V. et al. Axonal regeneration and functional recovery after complete spinal cord transection in rats by delayed treatment with transplants and neurotrophins. *J. Neurosci.* **21**, 9334–9344 (2001).
- Doyle, J. P. et al. Application of a translational profiling approach for the comparative analysis of CNS cell types. *Cell* **135**, 749–762 (2008).
- Lu, P. et al. Long-distance growth and connectivity of neural stem cells after severe spinal cord injury. *Cell* **150**, 1264–1273 (2012).
- Kumamaru, H. et al. Generation and post-injury integration of human spinal cord neural stem cells. *Nat. Methods* **15**, 723–731 (2018).
- Heiman, M., Kulicke, R., Fenster, R. J., Greengard, P. & Heintz, N. Cell type-specific mRNA purification by translating ribosome affinity purification (TRAP). *Nat. Protocols* **9**, 1282–1291 (2014).
- Groh, A. et al. Cell-type specific properties of pyramidal neurons in neocortex underlying a layout that is modifiable depending on the cortical area. *Cereb. Cortex* **20**, 826–836 (2010).
- Lein, E. S. et al. Genome-wide atlas of gene expression in the adult mouse brain. *Nature* **445**, 168–176 (2007).
- Lodato, S. et al. Gene co-regulation by *Fezf2* selects neurotransmitter identity and connectivity of corticospinal neurons. *Nat. Neurosci.* **17**, 1046–1054 (2014).
- Fink, K. L., Strittmatter, S. M. & Cafferty, W. B. Comprehensive corticospinal labeling with  $\mu$ -crystallin transgene reveals axon regeneration after spinal cord trauma in *ng2*<sup>−/−</sup> mice. *J. Neurosci.* **35**, 15403–15418 (2015).
- Woodworth, M. B., Greig, L. C., Kriegstein, A. R. & Macklis, J. D. SnapShot: cortical development. *Cell* **151**, 918–918.e1 (2012).
- Karow, M. et al. Reprogramming of pericyte-derived cells of the adult human brain into induced neuronal cells. *Cell Stem Cell* **11**, 471–476 (2012).
- Arlotta, P. et al. Neuronal subtype-specific genes that control corticospinal motor neuron development *in vivo*. *Neuron* **45**, 207–221 (2005).
- Hollis, E. R. II, Jamshidi, P., Löw, K., Blesch, A. & Tuszynski, M. H. Induction of corticospinal regeneration by lentiviral trkB-induced Erk activation. *Proc. Natl Acad. Sci. USA* **106**, 7215–7220 (2009).
- Park, K. K. et al. Promoting axon regeneration in the adult CNS by modulation of the PTEN/mTOR pathway. *Science* **322**, 963–966 (2008).
- Liu, Y. et al. A sensitized IGF1 treatment restores corticospinal axon-dependent functions. *Neuron* **95**, 817–833 (2017).
- Qiu, J. et al. Spinal axon regeneration induced by elevation of cyclic AMP. *Neuron* **34**, 895–903 (2002).
- He, Z. & Jin, Y. Intrinsic control of axon regeneration. *Neuron* **90**, 437–451 (2016).
- Sekine, Y., Siegel, C. S., Sekine-Konno, T., Cafferty, W. B. J. & Strittmatter, S. M. The nociceptin receptor inhibits axonal regeneration and recovery from spinal cord injury. *Sci. Signal.* **11**, eaao4180 (2018).
- Vodrazka, P. et al. The semaphorin 4D–plexin-B signalling complex regulates dendritic and axonal complexity in developing neurons via diverse pathways. *Eur. J. Neurosci.* **30**, 1193–1208 (2009).
- Hayashi, N., Miyata, S., Yamada, M., Kamei, K. & Oohira, A. Neuronal expression of the chondroitin sulfate proteoglycans receptor-type protein-tyrosine phosphatase  $\beta$  and phosphacan. *Neuroscience* **131**, 331–348 (2005).
- Maden, C. H. et al. NRP1 and NRP2 cooperate to regulate gangliogenesis, axon guidance and target innervation in the sympathetic nervous system. *Dev. Biol.* **369**, 277–285 (2012).
- Samuels, I. S., Saitta, S. C. & Landreth, G. E. MAPing CNS development and cognition: an ERKsome process. *Neuron* **61**, 160–167 (2009).
- Poplawski, G. et al. Adult rat myelin enhances axonal outgrowth from neural stem cells. *Sci. Transl. Med.* **10**, eaa12563 (2018).
- Dulin, J. N. et al. Injured adult motor and sensory axons regenerate into appropriate organotypic domains of neural progenitor grafts. *Nat. Commun.* **9**, 84 (2018).
- Adler, A. F., Lee-Kubli, C., Kumamaru, H., Kadoya, K. & Tuszynski, M. H. Comprehensive monosynaptic rabies virus mapping of host connectivity with neural progenitor grafts after spinal cord injury. *Stem Cell Reports* **8**, 1525–1533 (2017).
- Koffler, J. et al. Biomimetic 3D-printed scaffolds for spinal cord injury repair. *Nat. Med.* **25**, 263–269 (2019).
- Belin, S. et al. Injury-induced decline of intrinsic regenerative ability revealed by quantitative proteomics. *Neuron* **86**, 1000–1014 (2015).
- McKinstry, S. U. et al. Huntingtin is required for normal excitatory synapse development in cortical and striatal circuits. *J. Neurosci.* **34**, 9455–9472 (2014).
- Phillips, O. et al. The corticospinal tract in Huntington's disease. *Cereb. Cortex* **25**, 2670–2682 (2015).
- Zeitlin, S., Liu, J. P., Chapman, D. L., Papaioannou, V. E. & Efstratiadis, A. Increased apoptosis and early embryonic lethality in mice nullizygous for the Huntington's disease gene homologue. *Nat. Genet.* **11**, 155–163 (1995).
- Dietrich, P., Johnson, I. M., Alli, S. & Dragatsis, I. Elimination of huntingtin in the adult mouse leads to progressive behavioral deficits, bilateral thalamic calcification, and altered brain iron homeostasis. *PLoS Genet.* **13**, e1006846 (2017).
- Wang, G., Liu, X., Gaertig, M. A., Li, S. & Li, X. J. Ablation of huntingtin in adult neurons is nondeleterious but its depletion in young mice causes acute pancreatitis. *Proc. Natl Acad. Sci. USA* **113**, 3359–3364 (2016).
- Landles, C. & Bates, G. P. Huntingtin and the molecular pathogenesis of Huntington's disease. *EMBO Rep.* **5**, 958–963 (2004).
- Butler, R. & Bates, G. P. Histone deacetylase inhibitors as therapeutics for polyglutamine disorders. *Nat. Rev. Neurosci.* **7**, 784–796 (2006).
- Duyao, M. P. et al. Inactivation of the mouse Huntington's disease gene homolog *Hdh*. *Science* **269**, 407–410 (1995).
- Ross, C. A. et al. Huntington disease: natural history, biomarkers and prospects for therapeutics. *Nat. Rev. Neurol.* **10**, 204–216 (2014).
- Marcora, E. & Kennedy, M. B. The Huntington's disease mutation impairs Huntington's role in the transport of NF- $\kappa$ B from the synapse to the nucleus. *Hum. Mol. Genet.* **19**, 4373–4384 (2010).
- Trushina, E. et al. Mutant huntingtin impairs axonal trafficking in mammalian neurons *in vivo* and *in vitro*. *Mol. Cell. Biol.* **24**, 8195–8209 (2004).
- Dragatsis, I., Levine, M. S. & Zeitlin, S. Inactivation of *Hdh* in the brain and testis results in progressive neurodegeneration and sterility in mice. *Nat. Genet.* **26**, 300–306 (2000).
- Fouad, K., Pedersen, V., Schwab, M. E. & Brösamle, C. Cervical sprouting of corticospinal fibers after thoracic spinal cord injury accompanies shifts in evoked motor responses. *Curr. Biol.* **11**, 1766–1770 (2001).
- Weidner, N., Ner, A., Salimi, N. & Tuszynski, M. H. Spontaneous corticospinal axonal plasticity and functional recovery after adult central nervous system injury. *Proc. Natl Acad. Sci. USA* **98**, 3513–3518 (2001).
- Bareyre, F. M. et al. The injured spinal cord spontaneously forms a new intraspinal circuit in adult rats. *Nat. Neurosci.* **7**, 269–277 (2004).
- Patel, A. et al. AxonTracer: a novel ImageJ plugin for automated quantification of axon regeneration in spinal cord tissue. *BMC Neurosci.* **19**, 8 (2018).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2020

## Methods

No statistical methods were used to predetermine sample size.

### Study design

We examined the regeneration transcriptome of the corticospinal neuron triggered by growth-promoting NPC grafts compared to ungrafted, lesioned controls. We used mice with layer-5-neuron-specific expression of eGFP-L10a under the Glt25d2 promoter. We pooled three mice for mRNA collecting per condition for each of three time points plus three mice for a bone-marrow stromal cell (MSC) control-graft condition. For in vivo studies in mice, we used a sample size of 13 for *Htt*-wild-type mice and 15 for *Htt*-knockout mice. Numbers of replicates are listed in each figure legend. All experiments were randomized with regard to mouse enrolment into treatment group, and experimenters were blinded to experimental condition when performing data analyses.

### Mice

All procedures involving mice were carried out in strict adherence to guidelines provided by The Guide for the Care and Use of Laboratory Animals, and requirements of the institutional animal welfare committee. The Glt25d2 bacterial artificial chromosome (BAC) with eGFP-tagged ribosomal protein for TRAP (Glt25d2-eGFP-L10a) mice were obtained from the GENESAT project and maintained as previously described<sup>8</sup>. Experiments were performed on mice with one Glt25d2-eGFP-L10a and one wild-type C57Bl/6 parent. Glt25d2-eGFP-L10a mice were backcrossed ten times to the wild-type C57Bl/6 background to avoid rejection of grafted NPCs, derived from either C57Bl/6-Tg(CAG-GFP)10sb/J mice or from wild-type C57Bl/6 mice. C57Bl/6-Tg(CAG-GFP)10sb/J (eGFP mice) expressing eGFP cDNA under the control of a chicken  $\beta$ -actin promoter and cytomegalovirus enhancer were bred with C57Bl/6 (wild-type mice) female mice to produce eGFP-positive embryos. eGFP-mice and wild-type mice were obtained from Jackson Laboratory. Floxed *Htt* mice were generated by I.D. (ref. <sup>45</sup>) and backcrossed ten times to the wild-type C57Bl/6 background to avoid rejection of grafted NPCs. All mice were maintained on a 12-h light/dark cycle and given ad libitum access to food and water.

### Mouse surgery

All surgery was done under deep anaesthesia using a combination of ketamine (80–100 mg/kg) and xylazine (5–10 mg/kg). Euthanasia for tissue collection was performed by injection of an overdose amount of anaesthesia cocktail in accordance with AVMA Guidelines for Euthanasia of Laboratory Animals.

### Dorsal column lesion

To assess the regeneration of corticospinal tract axons, subjects underwent C2, C5 or thoracic (T)1 dorsal column lesions as previously described<sup>47</sup>. In brief, a Kopf wire knife (David Kopf Instruments) was inserted 0.3 mm lateral to midline and 1.1 mm under the dorsal surface of the spinal cord; the knife was extruded 1.5 mm and lifted to transect the dorsal columns, with coincident compression with a 28-ga blunt tip from above to ensure lesion completeness.

### Transplantation surgeries

For mouse-donor studies, grafting was performed 1 week after T1 dorsal column wire-knife lesion. E12 spinal cords from transgenic eGFP mice were dissected in ice-cold HBSS, digested in 0.25% trypsin for 20 min at 37 °C, washed 3 times with 10% FBS in neurobasal medium, triturated in 5 ml neurobasal medium + B27, filtered with 70- $\mu$ m mesh, spun at 200g for 5 min at room temperature, resuspended in 2 ml neurobasal medium + B27 and kept on ice. Before grafting, cells were spun at 200g for 3 min at room temperature, supernatant was aspirated and cell pellets were resuspended at a concentration of 200,000 cells per  $\mu$ l in a fibrin matrix (25 mg/ml fibrinogen and 25 U/ml thrombin, Sigma F6755

and T5772) containing growth factors to support graft survival<sup>1</sup>: BDNF (50  $\mu$ g/ml, Peprotech, 452-02), neurotrophin 3 (NT-3; 50  $\mu$ g/ml, Peprotech, 450-03), platelet-derived growth factor (PDGF-AA; 10  $\mu$ g/ml, Sigma, P3076), insulin-like growth factor 1 (IGF-1; 10  $\mu$ g/ml, Sigma, I8779), epidermal growth factor (EGF; 10  $\mu$ g/ml, Sigma, E1257), basic fibroblast growth factor (bFGF; 10  $\mu$ g/ml, Sigma, F0291), acidic fibroblast growth factor (aFGF; 10  $\mu$ g/ml, Sigma, F5542), glial-cell-line-derived neurotrophic factor (GDNF; 10  $\mu$ g/ml, Sigma, G1401), hepatocyte growth factor (HGF; 10  $\mu$ g/ml, Sigma, H9661) and calpain inhibitor (MDL28170, 50  $\mu$ M, Sigma, M6690). Thrombin and fibrinogen cell suspensions were injected separately to allow for fibrin matrix formation within the spinal cord. Fibrinogen suspension was injected in all sites before thrombin suspension. The graft mixture was microinjected into the lesion cavity at T1 and into the intact spinal cord at C4 and C6 in two locations each. Each injection site received 0.83  $\mu$ l total cell suspension, and subjects survived for either 3, 7 or 14 days after grafting. Lesion-only mice received growth factors in fibrinogen–thrombin mixture without cells.  $n = 24$  mice survived 3 days (9 with NPC graft, 9 without graft and 6 for histology),  $n = 24$  mice survived 7 days (9 with NPC graft, 9 without graft and 6 for histology) and  $n = 33$  mice survived 14 days after grafting (9 with NPC graft, 9 with MSC graft, 9 without graft and 6 for histology).

### Anterograde corticospinal tracing

Nine days before perfusion, corticospinal tract axons were anterogradely labelled by injection of 0.3  $\mu$ l of 10% BDA (MW 10,000, Molecular Probes) into each of 10 sites per hemisphere spanning the motor cortex (from bregma, rostral–caudal:  $\pm 0.0$ , 0.5, 1 mm; lateral:  $\pm 1.2$ , 2.2 mm; depth: 0.7 mm).

### Retrograde corticospinal tracing

To colabel CST neurons projecting into the cervical spinal cord that also express the BAC–eGFP–L10a transgene, we injected 0.2% cholera toxin B (CTB) in PBS bilaterally into the cervical (C4–C8) grey matter of Glt25d2-GFPL10a mice ( $n = 2$ ). Each mouse received a total of 8 injections of CTB at a volume of 300 nl per site, spanning 4 sites rostrocaudally ( $-0.5$  and  $+0.5$  mm lateral from midline). Mice were perfused three days after injection and cortices were processed for histological assessment.

### *Htt*-knockout experiments

AAV8 vectors expressing tdTomato and Cre recombinase (AAV8-tdTomato-Cre) were injected into the right motor cortices of four-month-old *Htt*<sup>flax/flax</sup> mice<sup>45</sup> ( $n = 15$ ) (Fig. 4), a late time point that would not affect neuronal development<sup>38</sup>. Six weeks later, we performed C4 spinal-cord dorsal-column lesions to transect the corticospinal projection, and grafted eGFP-expressing wild-type-mouse E12 spinal-cord-derived multipotent NPCs from the same mouse background into the lesion site; we hypothesized that corticospinal regeneration would be reduced in these mice owing to *Htt* deletion. A separate group of *Htt*<sup>flax/flax</sup> mice ( $n = 13$ ) received motor cortex injections of AAV8 vectors expressing tdTomato but not Cre, followed 6 weeks later by grafts of the same eGFP-expressing wild-type mouse E12 spinal-cord-derived multipotent NPC grafts to C4 spinal-cord dorsal-column lesions; we predicted that this group would exhibit corticospinal regeneration, as previously reported<sup>1</sup>. AAV8 viral vectors (Salk Institute Viral Vector Core) expressing either TdTomato (*Htt*<sup>+/+</sup> mice,  $n = 13$ ) or TdTomato plus Cre recombinase (*Htt*<sup>-/-</sup> mice,  $n = 15$ ) transgenes under control of the CAG promoter were injected into forelimb and hindlimb areas of the right primary motor cortex and primary somatosensory cortex of *Htt*<sup>flax/flax</sup> mice at a concentration of  $5 \times 10^{12}$  viral particles per ml, volume 0.5  $\mu$ l per site, 6 sites per side of the brain<sup>45</sup> (coordinates from bregma: anterior–posterior 1 to  $-1$  mm; medial–lateral 1.5 to 2.5 mm; depth 0.7 mm). Six weeks later, mice underwent a dorsal column lesion at C4 followed by immediate grafts of eGFP-expressing fetal spinal-cord-derived NPCs. Four weeks after grafting, mice were processed for histology. ( $n = 15$  *Htt* knockout;



# Article

*n* = 13 *Htt* wild type). Regenerating axons were quantified via the freely available ImageJ plugin, AxonTracer<sup>49</sup>.

## Fluorescent labelling

Spinal cords or brains were removed from the vertebrae and skull and serially sectioned in the sagittal or coronal plane at 35- $\mu$ m intervals. Every sixth section was used for antibody labelling. All steps were performed at room temperature unless noted otherwise. Sections were washed 3 times in TBST (0.025% triton-X100 in TBS) for 10 min, washed with 50% methanol in TBS for 5 min, dehydrated in 100% methanol for 15 min, washed with 50% methanol in TBS for 5 min and blocked in TBST + 5% donkey serum for 1 h, and then incubated overnight in TBST + 3% donkey serum with primary antibodies at 4 °C. Sections were washed 3 times in TBST + 3% horse serum and incubated 2.5 h with secondary antibodies, then washed 3 times with TBS for 10 min, mounted on glass slides and coverslipped Fluoromount G (Southern Biotechnology).

## Primary and secondary antibodies for fluorescent immunohistochemistry

Rabbit anti-GFP (1:1000, A-11122, Invitrogen) with Alexa-488-conjugated donkey anti-rabbit (1:200; Jackson ImmunoResearch); rabbit anti-RFP-biotin (1:5,000; ab34771; Abcam) with Alexa-594-conjugated streptavidin (1:300); mouse anti-NeuN (1:250, MAB377, Millipore) with Alexa-647-conjugated donkey anti-mouse (1:200; Jackson ImmunoResearch) for Fig. 3 and with Alexa-594-conjugated donkey anti-mouse for Supplementary Fig. 1; mouse anti-CTB (1:1,000, C86204M, Meridian Life Sciences) with Alexa-594-conjugated donkey anti-mouse; goat anti-SOX9 (1:1,000, AF3075, R&D Systems) and Alexa-594-conjugated donkey anti-goat; rabbit anti-PAX2 (1:500, 71600, Life Technology) and Alexa-594-conjugated donkey anti-rabbit; rabbit anti-ATG7 (1:500, LS-c156613, LS Bio) with Alexa-594-conjugated donkey anti-rabbit; rabbit anti-Cre (1:1,000; 908001, Biolegend) with Alexa-488-conjugated donkey anti-rabbit; nuclei were stained with 4',6-diamidino-2-phenylindole dihydrochloride (DAPI) at 1  $\mu$ g/ml (Sigma-Aldrich).

## Light-level HTT labelling

PFA fixed cortical sections were incubated in sodium citrate buffer (10mM sodium citrate, 0.05% Tween 20, pH 6.0) for 10 min at 90 °C; washed twice for 5 min with TBS; quenched with 0.06% H<sub>2</sub>O<sub>2</sub> in TBS for 20 min; washed twice for 2 min in TBS; blocked for 1 h (M.O.M. mouse IgG blocking reagent in 1 ml TBS), washed twice for 2 min in TBS; incubated in diluent for 5 min (400  $\mu$ l of protein concentrate stock solution in 5 ml TBS); incubated with mouse HTT (1:100) MAB2166 in diluent overnight at room temperature, washed 3 times for 10 min each in TBS incubated with M.O.M. biotinylated anti-mouse IgG reagent for 2.5 h; washed 3 times for 10 min each in TBS; incubated in VECTASTAIN Elite ABC HRP Kit (Vector Laboratories) for 30 min; washed 3 times for 10 min each in TBS; incubated in biotinyl tyramide (1:2,500) in 30% H<sub>2</sub>O<sub>2</sub> in TBS for 30 min.; washed 3 times for 10 min each in TBS; incubated with ABC (9  $\mu$ l + 9  $\mu$ l in PBS) for 1 h; washed 3 times with TBS for 10 min each and stained with DAB; dehydrated with increasing concentrations of ethanol, followed by isopropanol and citrisolve for 2 min each.

## RNA isolation and sequencing

We performed dissections of the motor cortex in a total of 63 Glt25d2-eGFP-L10a mice (3 mice pooled per sample). Three intact samples (pooled from 9 mice) were studied, and a total of 18 lesioned samples (pooled from 54 mice) were studied at time points of 10, 14 and 21 days after injury. Experimental time lines and procedures are illustrated in Fig. 1a, b and Extended Data Fig. 1e. Motor cortices were dissected, eGFP-tagged polyribosomes were affinity-purified via eGFP-antibody-protein-L complexes using streptavidin-labelled magnetic beads, and separate bound and unbound RNA fractions (consisting of CST neurons and non-CST material, respectively) were isolated<sup>12,50</sup>. RNA was collected using the Absolutely RNA Nanoprep Kit (Agilent)

according to the manufacturer's instructions. Samples were shipped on dry ice to the University of California Los Angeles RNA-seq core for RNA quality control and RNA-seq on an Illumina HiSeq 2500 sequencer. Total RNA integrity was examined using the Agilent Bioanalyzer 2000 (Agilent) and quantified with NanoDrop (Thermo Fisher Scientific). All but one sample passed RNA quality assessment; one sample from the lesion control group at 10 days after injury was excluded. Total RNA (20 ng) was used to generate cDNA using Ovation RNA-Seq System V2 (NuGEN) following the manufacturer's instructions. cDNA (100 ng) was used in the library preparation using Ovation Ultralow Library Systems (NuGEN). The cDNA was fragmented to 300 base pairs (bp) using a Covaris M220 (Covaris), and then the manufacturer's instruction was followed for end repair, adaptor ligation and library amplification. All samples were multiplexed into a single pool to avoid batch effects<sup>51</sup> and sequenced using an Illumina HiSeq 2500 sequencer (Illumina) across 4 lanes of 69-bp paired-end sequencing, corresponding to 3 samples per lane and yielding between 31 and 69 (average of 42) million reads per sample. Alignment to the *Mus musculus* (mm10) refSeq (refFlat) reference gene annotation was performed using the STAR-spliced read aligner (v.2.3.0)<sup>52</sup> with default parameters. Data quality was assessed on base-quality calls, nucleotide composition of sequences, insert sizes, per cent of uniquely aligned reads and transcript coverage using custom scripts (available at <https://github.com/icnn/RNAseq-PIPELINE>) and picard-tools-1.118 (<http://broadinstitute.github.io/picard/>). Read alignments were visualized using the Integrative Genomics Viewer<sup>53</sup>. Sixty-four to seventy-seven per cent (average of 71%) of the reads mapped uniquely to the mouse genome. Total counts of read fragments aligned to candidate gene regions were derived using the HTSeq program (<https://htseq.readthedocs.io>) with mouse mm10 (Dec.2011) refSeq (refFlat table) as a reference and used as a basis for the quantification of gene expression. Only uniquely mapped reads were used for subsequent analyses. Differential expression analysis was conducted with R-project and the Bioconductor package edgeR (v.3.14.0)<sup>54</sup>. Statistical significance of differential expression was determined at FDR < 10% (*q* < 0.1). Ingenuity Pathway Analysis software (Qiagen) was used for data analysis (the complete gene list was used).

## Correlation with a previously published developmental dataset

We applied ranked gene expression by log-transformed fold change to each dataset individually. This was derived by comparing our non-regenerating and regenerating data to our intact dataset, and by comparing embryonic day 18 (E18) CST neurons to postnatal day 14 (P14) CST neurons in the previously published dataset<sup>18</sup>. We then correlated both datasets via rank-rank hypergeometric overlap<sup>55</sup>.

## Statistical analysis

All bar graphs are presented as the mean + s.e.m. For quantification, an unpaired, two-tailed Student's *t*-test was used. For RNA-seq analysis, an FDR of 10% (FDR *q* < 0.1) was applied; a statistical significance criterion of \**P* ≤ 0.05, \*\**P* ≤ 0.01, \*\*\**P* ≤ 0.001, \*\*\*\**P* ≤ 0.0001 was used in all other experiments. For Fig. 1d, one-way ANOVA with Tukey post hoc test was used; for Fig. 2d, Extended Data Fig. 9, we applied ranked gene expression by log-transformed fold change to both datasets individually. Experimenters were blinded to group identity when conducting experiments and performing analyses.

## Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

## Data availability

RNA-seq data have been deposited in the Gene Expression Omnibus (GEO) repository ([www.ncbi.nlm.nih.gov/geo/](http://www.ncbi.nlm.nih.gov/geo/)), accession number GSE126957.

50. Heiman, M. et al. A translational profiling approach for the molecular characterization of CNS cell types. *Cell* **135**, 738–748 (2008).
51. Auer, P. L. & Doerge, R. W. Statistical design and analysis of RNA sequencing data. *Genetics* **185**, 405–416 (2010).
52. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
53. Robinson, J. T. et al. Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–26 (2011).
54. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
55. Plaisier, S. B., Taschereau, R., Wong, J. A. & Graeber, T. G. Rank–rank hypergeometric overlap: identification of statistically significant overlap between gene-expression signatures. *Nucleic Acids Res.* **38**, e169 (2010).

**Acknowledgements** This work was supported by the Dr. Miriam and Sheldon G. Adelson Medical Research Foundation, the Veterans Administration (The Gordon Mansfield Consortium for Spinal Cord Regeneration), the NIH (NS09881 and EB014986) and the Gerbic Family Foundation. We acknowledge the support of the NINDS Informatics Center for Neurogenetics and Neurogenomics (P30 NS062691 to G.C.). We appreciate the assistance of J. Dulin.

**Author contributions** G.H.D.P. and M.H.T. conceptualized the study; G.H.D.P., R.K., G.C. and M.H.T. performed analysis of transcriptomic data; G.H.D.P., P.C. and R.L. performed data validation; G.H.D.P., R.L., N.M. and P.C. performed formal analysis; G.H.D.P., R.L., N.M., P.C., E.V.N. and J.M.M. performed investigation; G.H.D.P., N.M., P.C. and R.L. performed experiments and analysis for Fig.1; G.H.D.P., R.K. and G.C. performed experiments and data analysis and validation for Fig. 2, Extended Data Figs. 2, 6, 7; E.V.N. and J.M.M. performed experiments and analysis for Fig. 3; P.L. performed experiments for Extended Data Fig. 1; G.H.D.P. performed analysis for Extended Data Fig. 3; G.H.D.P. performed experiments and analysis for Extended Data Fig. 5; G.H.D.P. performed experiment and analysis for Extended Data Fig. 7; I.D., B.Z., G.C. and M.H.T. provided materials used for experiments; G.H.D.P., R.K. and G.C. performed data curation; G.H.D.P. and M.H.T. wrote the original draft; G.H.D.P., M.H.T., R.K. and G.C. reviewed and edited the manuscript.

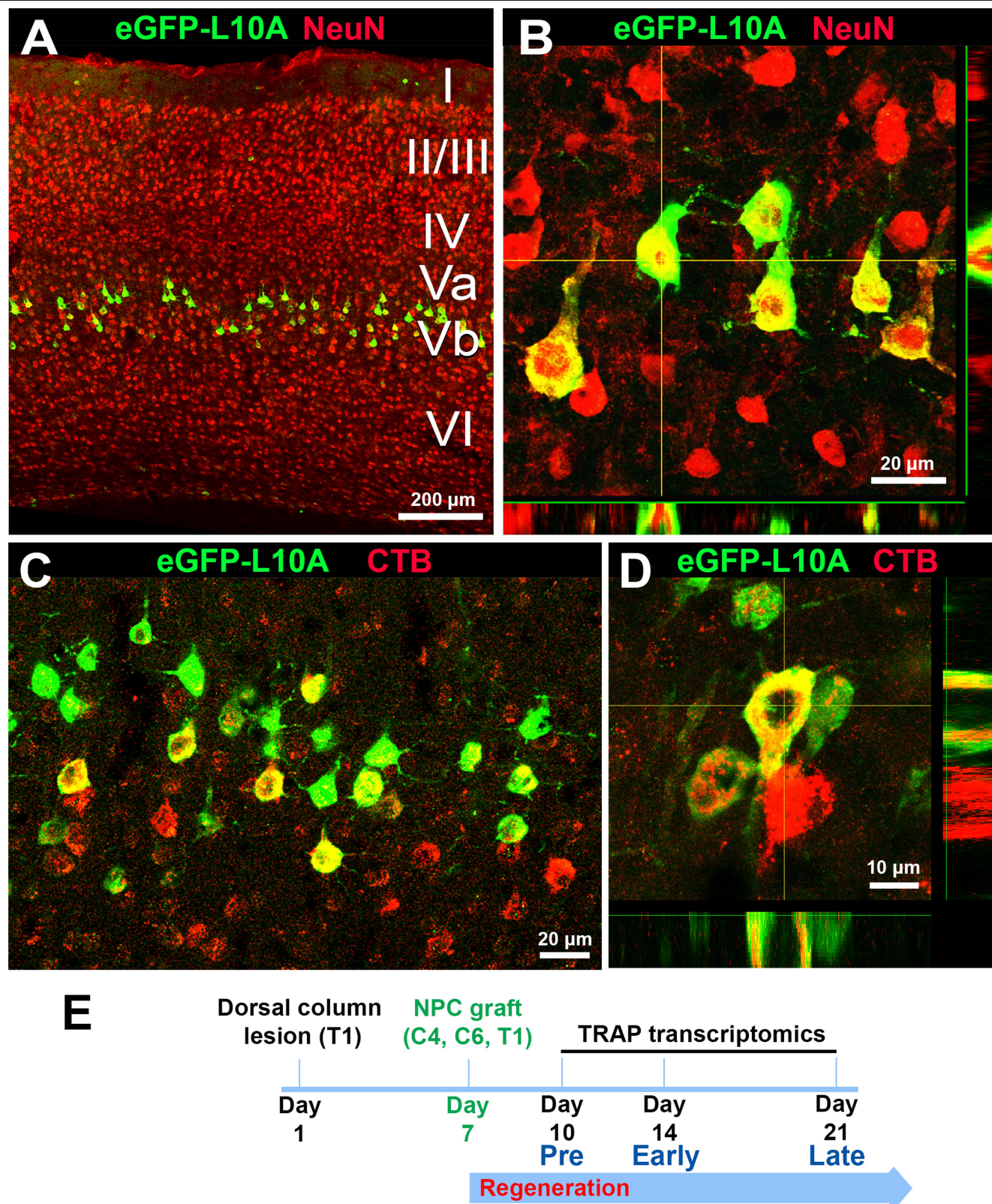
**Competing interests** The authors declare no competing interests.

#### Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41586-020-2200-5>.

**Correspondence and requests for materials** should be addressed to G.H.D.P. or M.H.T.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>.

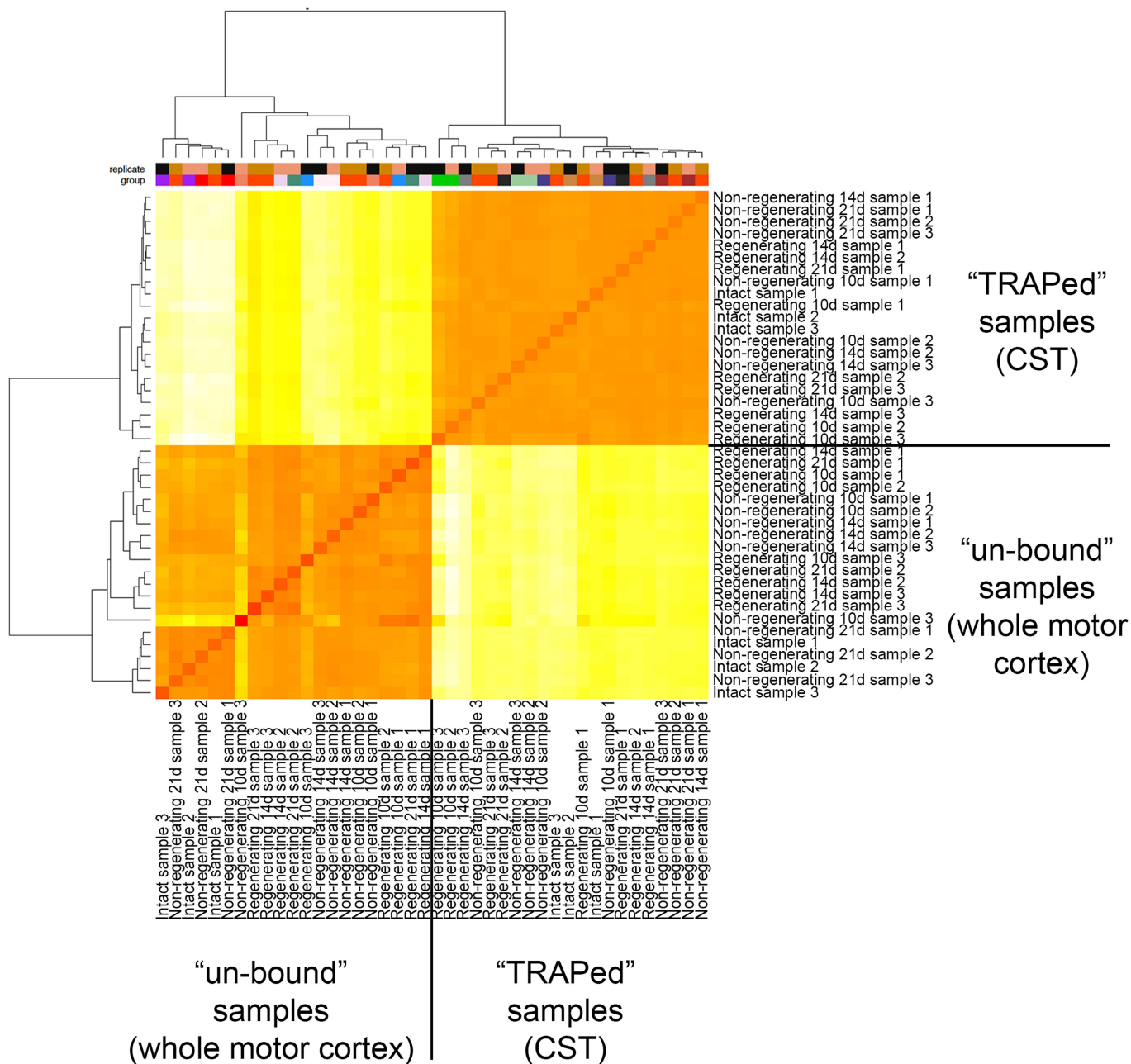


**Extended Data Fig. 1 | TRAP of regenerating corticospinal neurons.**

**a, b**, Coronal cortical section of transgenic BAC-TRAP mice expressing eGFP-L10A under the Glt25d2 promoter primarily in neurons (NeuN) of cortical layer 5b. **b**, Confocal z-stack projection of 15- $\mu$ m thickness showing colocalization of eGFP-L10A and NeuN within pyramidal neurons. **c, d**, Retrogradely labelled CST neurons (CTB) projecting to the cervical spinal cord (C4–C8) grey matter

colabel with eGFP-L10A, indicating that about 50% of TRAPed mRNA molecules are derived from corticospinal neurons. **d**, Confocal z-stack projection of 10- $\mu$ m thickness showing colocalization of eGFP-L10A and CTB within pyramidal neurons.  $n = 3$  biological replicates, which showed comparable results. **e**, Experimental timeline for purification of TRAPed mRNA molecules exclusively from cortical-layer-5 neurons.

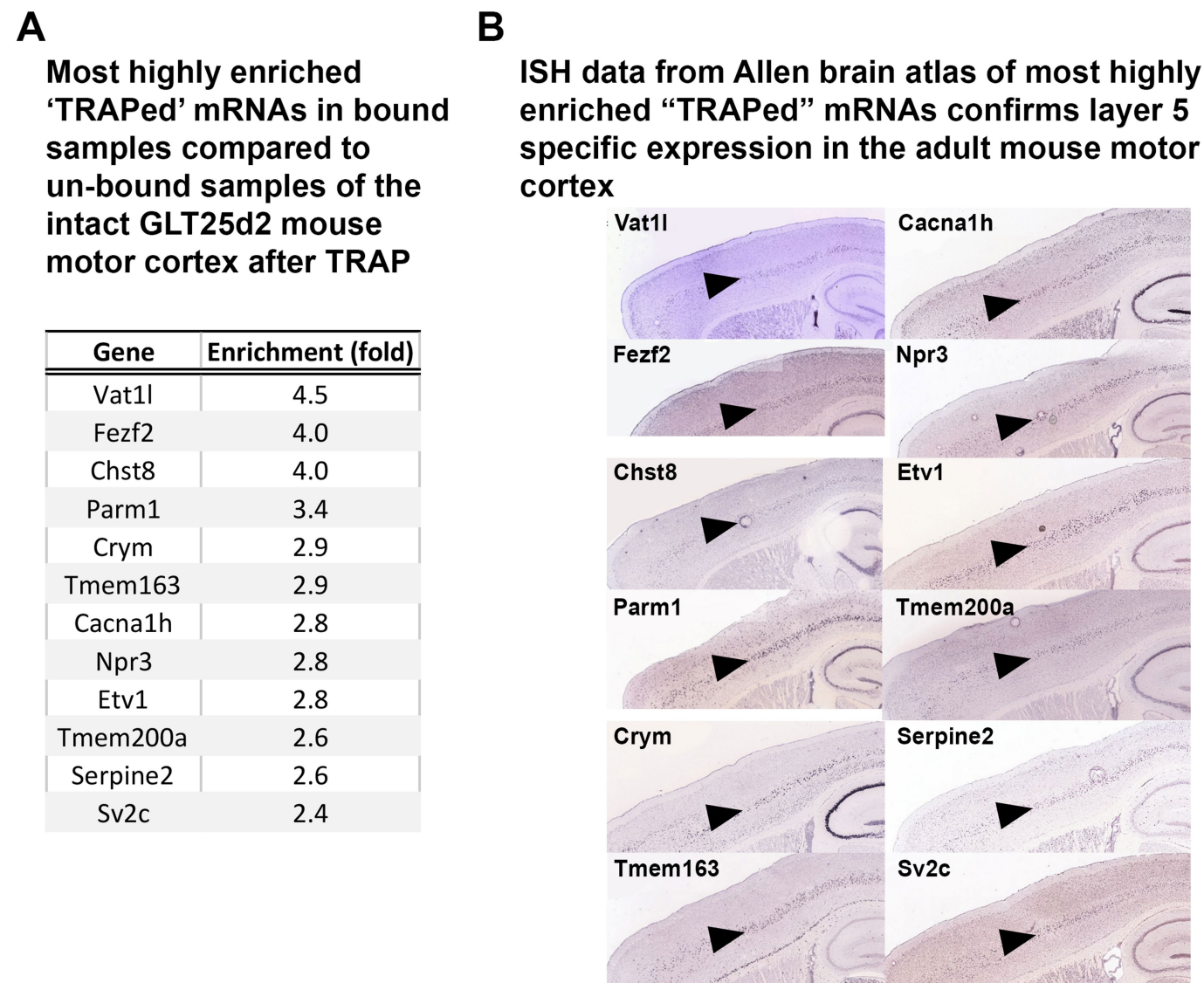




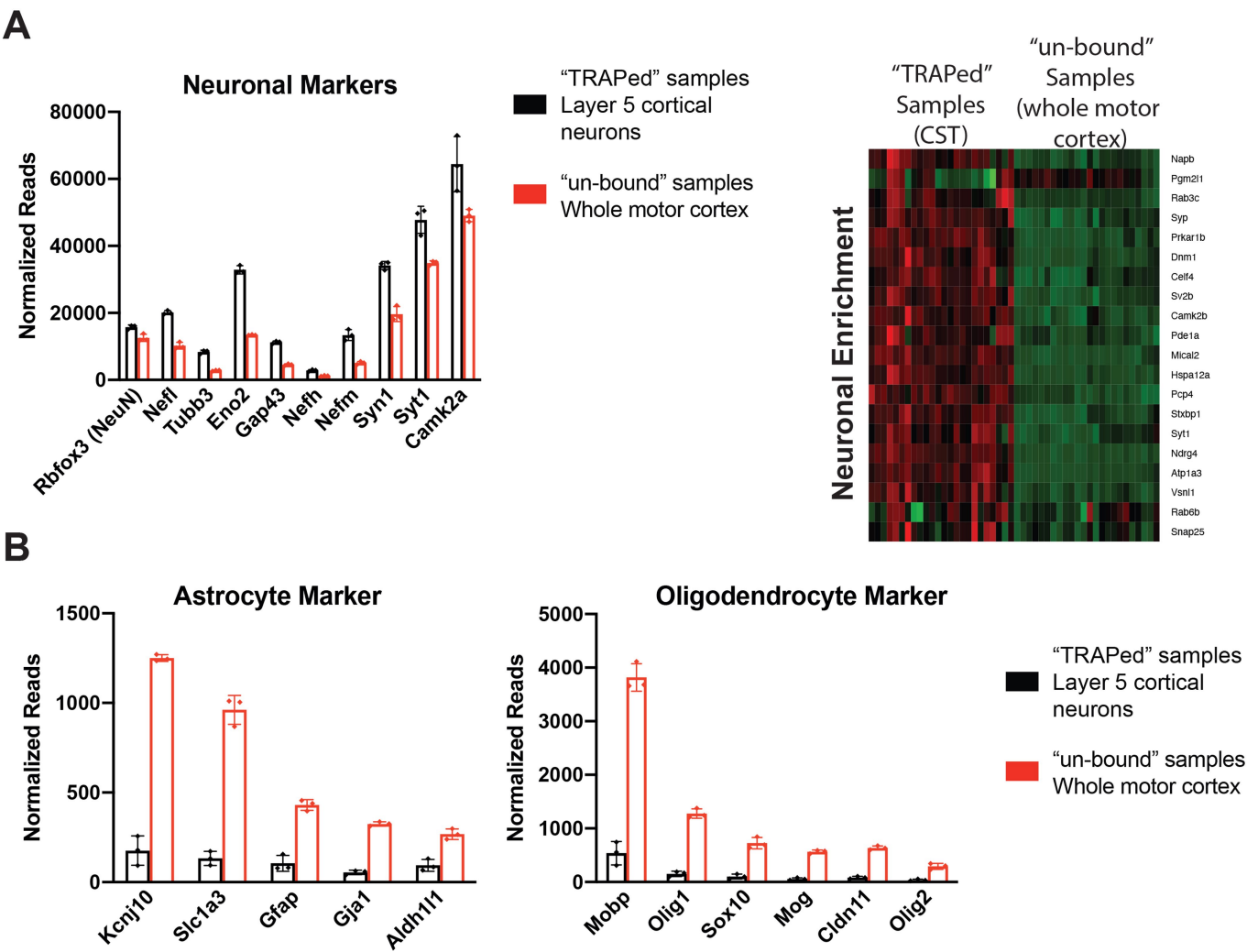
**Extended Data Fig. 2 | Sample Pearson correlation identifies unique populations of mRNAs derived from CST neurons (TRAPed samples) and from whole motor cortex (unbound samples) after TRAP.** mRNA species from all conditions are more similar to each other (TRAPed samples) than total

RNA collected from the motor cortex (unbound samples), indicating that TRAP enriched for a uniquely defined mRNA population that is distinct from whole cortex samples. Samples are arranged by hierarchical clustering.



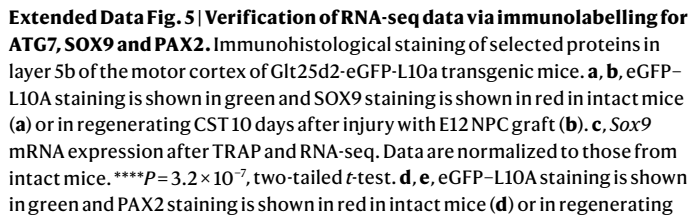


**Extended Data Fig. 3 | In situ hybridization data from the Allen Mouse Brain Atlas confirms layer-5b-specific mRNA expression of genes highly enriched by TRAP from intact Glt25d2-eGFP-L10a mice. a,** Fold change of most highly enriched mRNA transcripts in the corticospinal cell population (TRAPed samples) compared to mRNA from whole motor cortex (unbound samples) in the intact state. *n* = 3 biological replicates per condition. **b,** In situ hybridization (ISH) data (Allen Mouse Brain Atlas; <https://mouse.brain-map.org/>) of most highly enriched genes (from **a**) confirms cell-type-specific enrichment of TRAPed mRNA expression in layer-5b neurons of the adult mouse motor cortex.



**Extended Data Fig. 4 | Neuronal genes are enriched in the TRAPed samples from layer-5b neurons.** **a**, Neuronal marker genes are enriched in the TRAPed samples derived from layer-5b neurons. **b**, Astrocyte- and

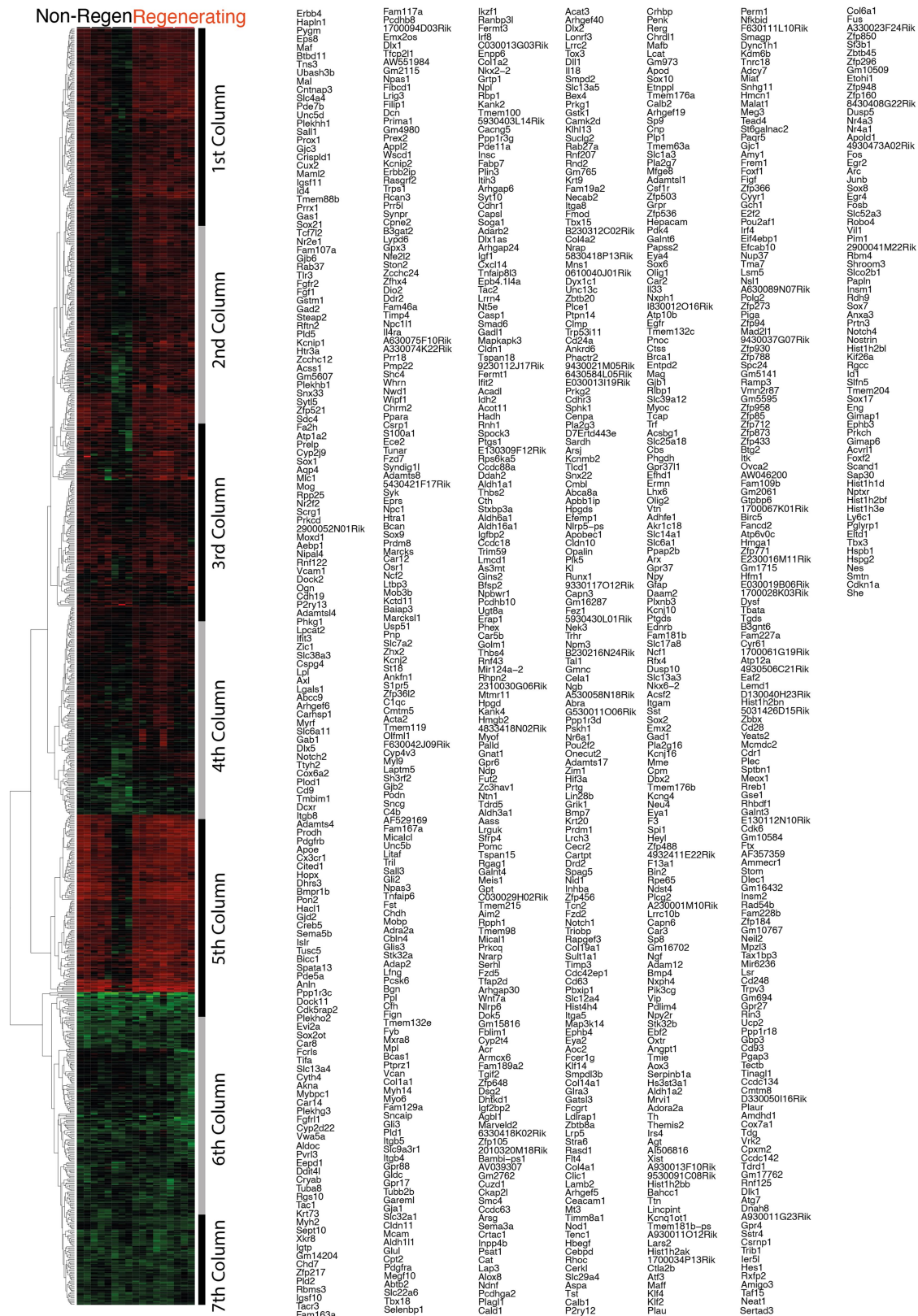
oligodendrocyte-specific markers are expressed at a higher level in the whole motor cortex (unbound samples).  $n = 3$  biological replicates per condition. Data are mean  $\pm$  s.d.



CST 10 days after injury with E12 NPC graft (**e**). **f**, *Pax2* mRNA expression after TRAP and RNA-seq. Data are normalized to those from intact mice. **\*\****P* = 0.0057, two-tailed *t*-test. **g**, **h**, eGFP–L10A staining is shown in green and ATG7 staining is shown in red in intact mice (**g**) or in regenerating CST 10 days after injury with E12 NPC graft (**h**). **i**, *Atg7* mRNA expression after TRAP and RNA-seq. Data are normalized to those from intact mice. **\*\*\*\****P* =  $3.0 \times 10^{-6}$ , two-tailed *t*-test. *n* = 3 biological replicates per condition. Data are mean  $\pm$  s.d.



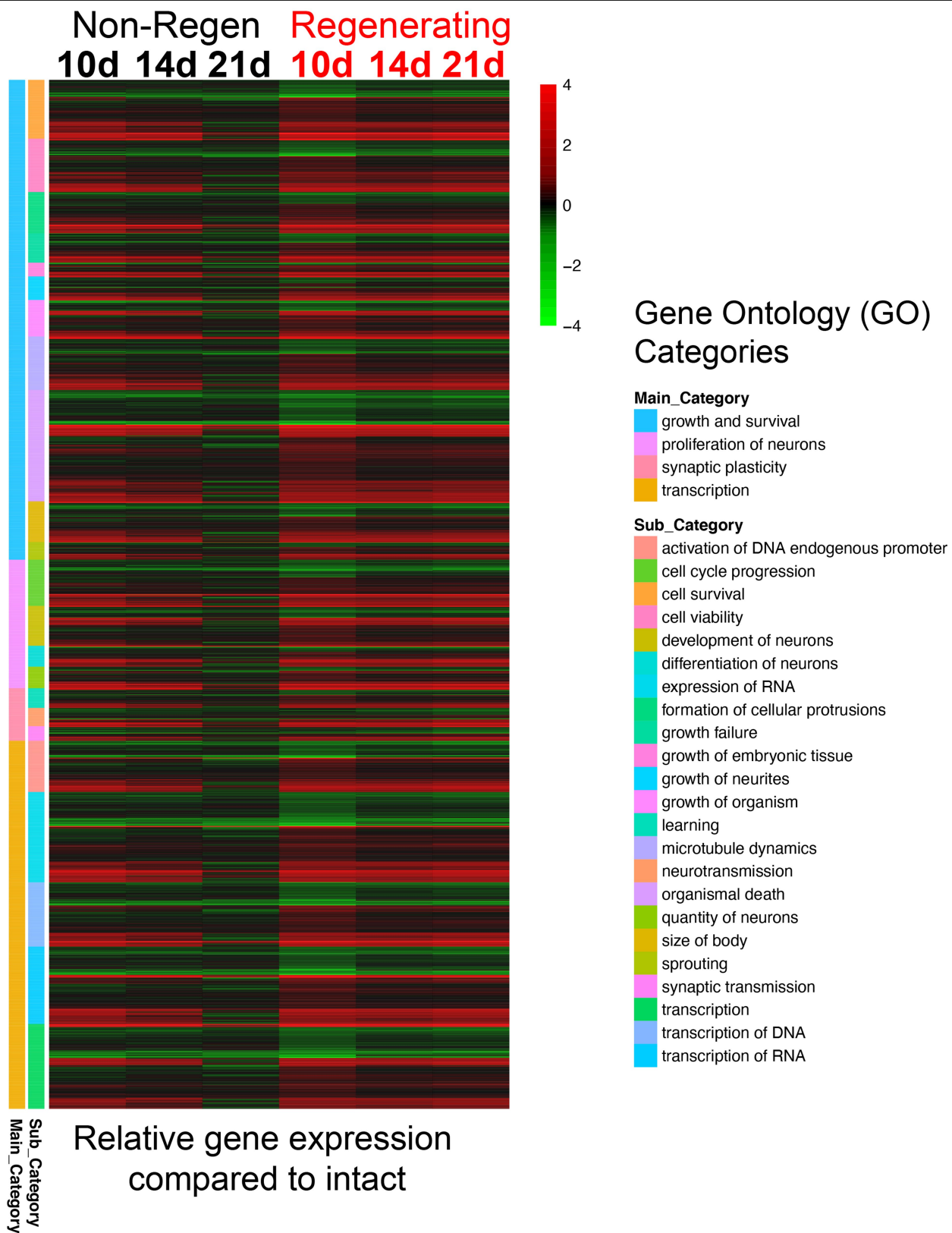
# Non-RegenRegenerating



**Extended Data Fig. 6 | Heat map of the top 1,000 significantly differentially regulated transcripts (FDR < 0.1) upon injury and upon regeneration, relative to intact mice, with individual gene names.** Heat map of the top 1,000 significantly differentially regulated transcripts (FDR < 0.1, Benjamini–Hochberg, two-sided) upon injury and upon regeneration, relative to the intact state. Red, increased expression; green, reduced expression. The

intensity of the colour reflects the degree of gene regulation compared to the intact state. Transcripts are arranged by hierarchical clustering,  $n = 2$  biological replicates for day-10 non-regenerating, and  $n = 3$  biological replicates for all other conditions. Genes are listed from top to bottom. Black and grey boxes next to heat map indicate gene lists (columns 1 to 7 from left to right).

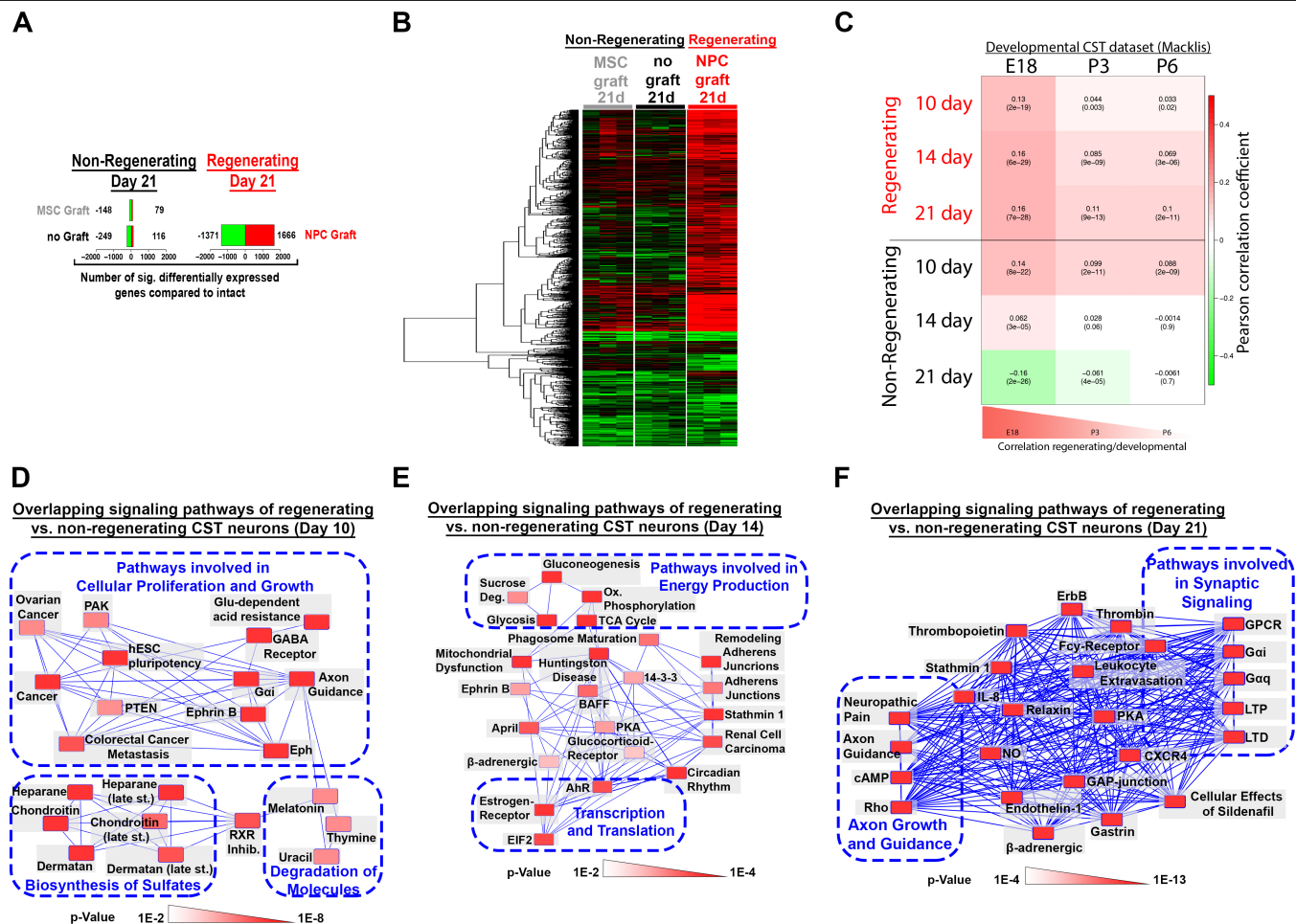




Extended Data Fig. 7 | See next page for caption.

**Extended Data Fig. 7 | Heat map of genes enriched in the Gene Ontology analysis (Ingenuity Pathway Analysis software).** Heat map of Gene Ontology (GO)-term-associated significantly differentially regulated transcripts (FDR < 0.1, Benjamini–Hochberg, two-sided) upon injury and upon regeneration, relative to the intact state, at 10, 14 and 21 days after injury. Red, increased expression; green, reduced expression. The intensity of the colour reflects the degree of gene regulation compared to the intact state. Transcripts are arranged by GO-term clustering (Ingenuity Pathway Analysis software). GO clusters listed on the right are indicated via colour code on the left side of the

heat map. Many common genes have increased in expression in both non-regenerating and regenerating groups 10 days after injury. However, by day 14, these injury-induced changes are diminishing in the absence of a graft but are sustained in the presence of a graft, into which regeneration occurs. By day 21 after injury, many genes have returned to baseline levels in non-regenerating CST neurons, whereas they remain highly upregulated in regenerating CST neurons. The GO categories are derived from Figs. 2, 3.  $n = 2$  biological replicates for day-10 non-regenerating, and  $n = 3$  biological replicates for all other conditions.

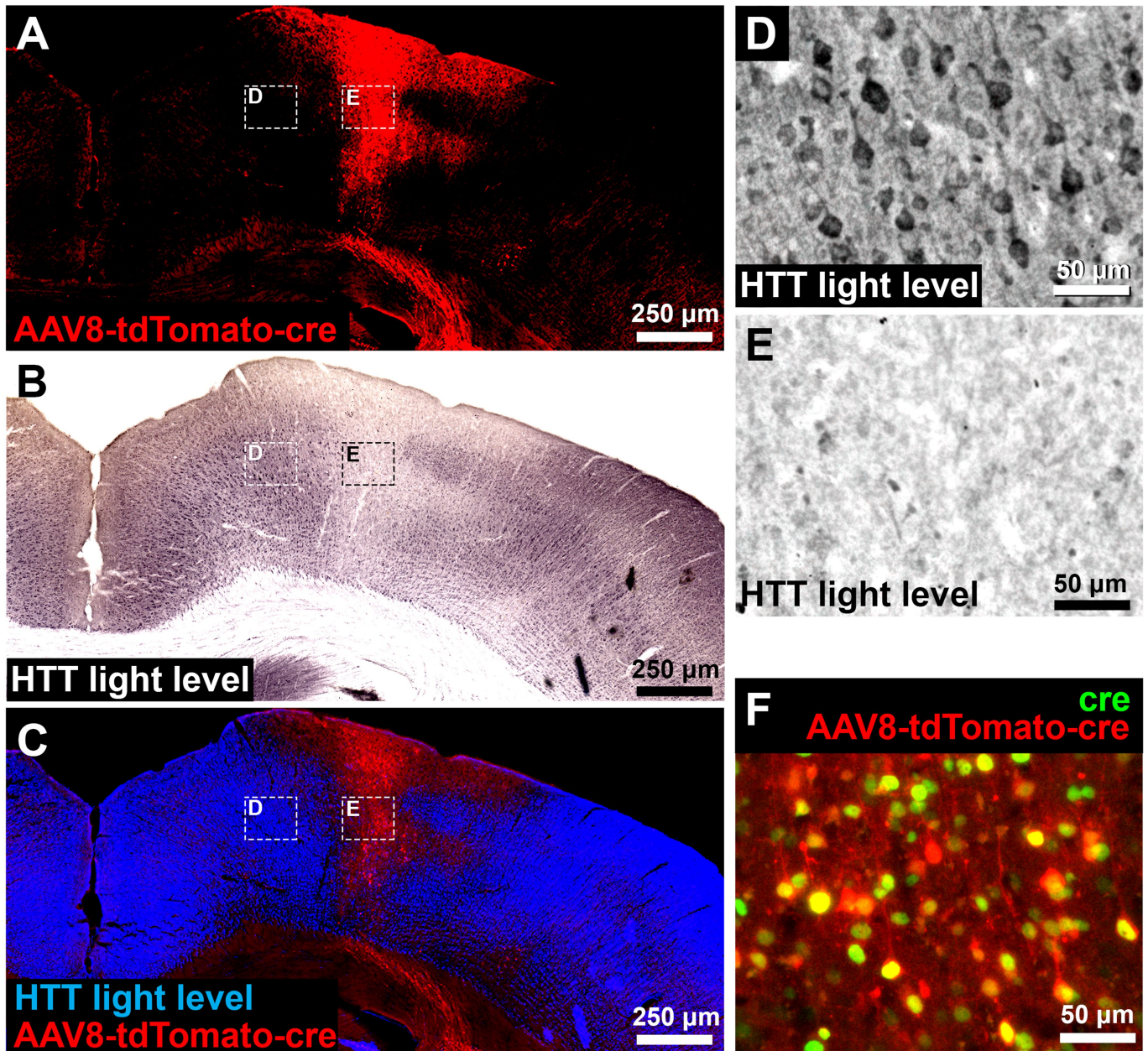


**Extended Data Fig. 8 | The regeneration gene-expression profile is sustained at 21 days only with an NPC graft, and not with a non-stimulatory MSC graft, and correlates with neuronal developmental transcriptomes.**

**a**, Number of significantly differentially expressed genes of non-regenerating (no graft and MSC graft) and regenerating (NPC graft) CST neurons relative to intact CST neurons (FDR < 0.1, Benjamini–Hochberg, two-sided) at 21 days after injury. Red, increased expression; green, reduced expression. **b**, Heat map of the top 1,000 significantly differentially regulated transcripts (FDR < 0.1) upon injury and upon regeneration, relative to the intact state, at 21 days after injury. Red, increased expression; green, reduced expression. The intensity of the colour reflects the degree of gene regulation compared to the intact state. Transcripts are arranged by hierarchical clustering. The non-regenerating conditions (no graft and MSC graft) show gene-expression patterns similar to one another; and, in both cases, gene expression is not sustained in comparison to the regenerating condition (NPC graft).  $n = 3$  biological replicates. **c**, Differentially expressed genes (FDR < 0.1) of a previous neuronal development study (GSE2039) for different time points against P14

(log-transformed fold change) was compared to CST data against the intact condition. Numbers indicate the Pearson correlation coefficient of the two datasets. Colour indicates the magnitude of the correlation coefficient (side bar). Values in parentheses are statistical significance ( $P$  values) of the correlation coefficients.  $n = 2$  biological replicates for day-10 non-regenerating and  $n = 3$  biological replicates for all other conditions. **d–f**, Overlapping signalling pathways of regenerating versus non-regenerating CST neurons (FDR < 0.1) using Ingenuity Pathway Analysis software at 10 (**d**), 14 (**e**) and 21 (**f**) days after injury. The network of overlapping signalling pathways shows each pathway as a single ‘node’ that is coloured proportionally to the right-tailed Fisher’s exact test  $P$  value; darker red, more significant. Individual scaling of  $P$  values is listed at the bottom of each graph. A line connects any two pathways when there is at least one dataset molecule in common between them. Pathways of similar function are clustered together (dashed blue outlines).  $n = 2$  biological replicates for day-10 non-regenerating and  $n = 3$  biological replicates for all other conditions.



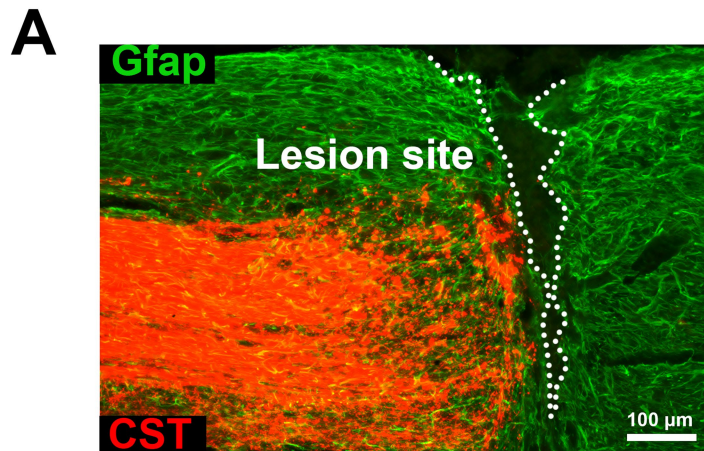


**Extended Data Fig. 9 | AAV8-tdTomato-Cre expression completely abolishes detection of HTT protein six weeks after injection.**

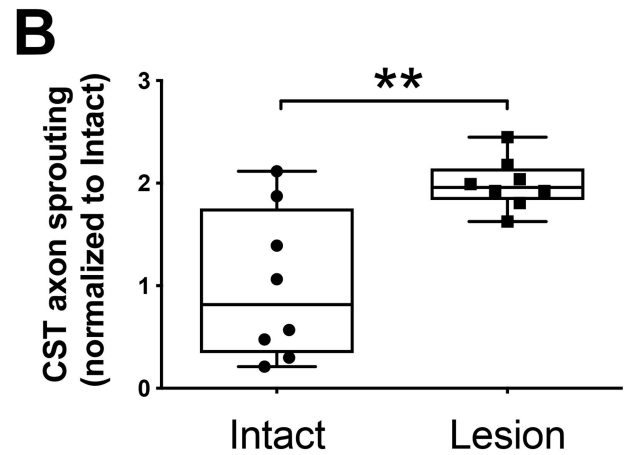
AAV8-dependent Cre delivery to *Htt<sup>flox/flox</sup>* mice was used to conditionally knock out HTT expression in corticospinal neurons six weeks before injury. **a–c**, Low magnification of right motor cortex six weeks after AAV8-tdTomato-Cre injection indicates successful tdTomato (red) expression in motor cortex and

reduction in HTT protein (black or blue) in the infected area. **d, e**, High-magnification images showing cortical neurons with baseline HTT protein expression outside of infected areas (**d**) and after HTT deletion in cortical neurons that are transfected with AAV8-tdTomato-Cre (**e**). **f**, AAV8-tdTomato-Cre-infected cortical neurons co-express tdTomato (red) and Cre (green).  $n = 3$  biological replicates, with comparable outcomes.





**Extended Data Fig. 10 | Lesioned corticospinal axons do not regenerate in the absence of an NPC graft. a,** CST axons labelled with the anterograde axonal tracer BDA do not regenerate in the absence of a growth-stimulating NPC graft. The GFAP label indicates glial activation around the lesion site. **b,** However, CST axons in mice without grafts do sprout into spared host grey



matter underlying the lesion, consistent with previous reports<sup>46–48</sup>. In the box plot, centre line is the median, box extends from the 25th to the 75th percentile and whiskers denote minimum–maximum values.  $n = 8$  biological replicates. **\*\*** $P = 0.0027$ , two-tailed, non-paired  $t$ -test.

## Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

### Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- ☐ ☒ The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement
- ☐ ☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- ☐ ☒ The statistical test(s) used AND whether they are one- or two-sided  
*Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
- ☐ ☒ A description of all covariates tested
- ☐ ☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- ☐ ☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- ☐ ☒ For null hypothesis testing, the test statistic (e.g.  $F$ ,  $t$ ,  $r$ ) with confidence intervals, effect sizes, degrees of freedom and  $P$  value noted  
*Give  $P$  values as exact values whenever suitable.*
- ☐ ☒ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- ☐ ☒ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- ☐ ☒ Estimates of effect sizes (e.g. Cohen's  $d$ , Pearson's  $r$ ), indicating how they were calculated

*Our web collection on [statistics for biologists](#) contains articles on many of the points above.*

### Software and code

Policy information about [availability of computer code](#)

Data collection

N.A.

Data analysis

ImageJ 1.8.0\_172, Excel 16.3.3 (Microsoft), Prism 8.3.0 (GraphPad), IPA (Qiagen)

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

### Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

RNAseq data has been deposited in the Gene Expression Omnibus (GEO) repository ([www.ncbi.nlm.nih.gov/geo](http://www.ncbi.nlm.nih.gov/geo)), accession number GSE126957. And will be accessible after the manuscript has been accepted.

### Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

- ☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

# Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	For TRAP analysis we combined n=3 mice per group to harvest enough mRNA in good quality for RNA-sequencing. Every time-point had n=3 biological replicates as is standard for RNAseq analysis. For HTT-KO experiments : A sample size of 11 subjects/group was determined by power analysis: 12 animals per group, standard deviation of 0.15, and effect size of 0.24 yields computed power of 0.85. This design is sufficient to determine significant group differences .We thus used at least n=15 subjects per group to account for potential issues such as incomplete graft integration.
Data exclusions	We excluded one replicate of the non-regeneration group at 10 days post lesion since the mRNA from this sample did not pass quality assessment. For the assessment of regeneration into the NPC graft we excluded 4 animals in the group that received AAV-TdTomato and 2 animal in the group that received AAV-TdTomato-Cre due to poor graft survival. The exclusion criteria was pre-established. The NPC graft has to fill the complete lesion cavity and cells have to differentiate to 80% into NeuN positive cells.
Replication	For TRAP analysis every experiment was performed in triplicates. all experiments were successful except 1 replicate of the non-regeneration group at 10 days post lesion did not pass quality assessment.
Randomization	Mice for TRAP were randomly assigned into groups to mix gender and age in every group of pooled animals for the different time-points investigated.
Blinding	Investigators were blinded by randomization to the analysis of HTT-floxed mice that received either AAV-cre-tdTomato or AAV-tdTomato

# Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems		Methods	
n/a	Involved in the study	n/a	Involved in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies	<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines	<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology	<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging
<input type="checkbox"/>	<input checked="" type="checkbox"/> Animals and other organisms		
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants		
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data		

## Antibodies

Antibodies used	rabbit anti-GFP (1:1000, A-11122, Invitrogen) with Alexa 488-conjugated donkey anti-rabbit (1:200; Jackson ImmunoResearch, West Grove, PA); rabbit anti-rfp-biotin (1:5000; ab34771; Abcam) with Alexa 594-conjugated Streptavidin (1:300); mouse anti-NeuN (1:250, MAB377, Millipore) with Alexa 647-conjugated donkey anti-mouse (1:200; Jackson ImmunoResearch, West Grove, PA) for Fig.3 and with Alexa 594-conjugated donkey anti-mouse for Fig.S1; mouse anti-CTB (1:1000, C86204M, Meridian Life Sciences) with Alexa 594-conjugated donkey anti-mouse; goat anti-Sox9 (1:1000, AF3075, R&D Systems) and Alexa 594-conjugated donkey anti-goat; rabbit anti-Pax2 (1:500, 71600, Life technology) and Alexa 594-conjugated donkey anti-rabbit; rabbit anti-ATG7 (1:500, LS-c156613, LS Bio) with Alexa 594-conjugated donkey anti-rabbit; rabbit anti-cre (1:1000; 908001, Biolegend) with with Alexa 488-conjugated donkey anti-rabbit; nuclei were stained with 4',6-diamidino-2-phenylindole dihydrochloride (DAPI; [1 µg/mL]; Sigma-Aldrich, St. Louis, MO). ms-Htt (1:100, AB2166, Millipore)
Validation	A-11122 validated by PMID: 30429769 for IHC Ab34771 validated by PMID: 30948479 for IHC MAB377 validated by PMID: 30948479 for IHC C86204M validated by PMID: 28361990 for IHC AF3075 validated by PMID: 28336567 for IHC 71600 validated by PMID: 30948479 for IHC C156613 validated by supplier for IHC 908001 validated by PMID: 27462817 for IHC MAB2166 validated by PMID: 21994396 for IHC

## Animals and other organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research

Laboratory animals	mice, C57Bl/6, mixed gender (age 6 to 12 month) mice, C57Bl/6-HTT-flox-flox, mixed gender (age 6 to 12 month) mice, C57Bl/6-Glt25d2-eGFP-L10a, mixed gender (age 6 to 12 month)
Wild animals	The study did not involve wild animals
Field-collected samples	The study did not contain field-collected samples
Ethics oversight	All procedures involving animals were carried out in strict adherence to guidelines provided by The Guide for the Care and Use of Laboratory Animals (The Institute of Laboratory Animal Resources, 2011), and requirements of the institutional animal welfare committee. National Institutes of Health guidelines for laboratory animal care and safety were strictly followed. All animal procedures were approved by the Institutional Animal Care and Use Committee of the Department of Veterans Affairs (VA) San Diego Healthcare System. Animals had free access to food and water throughout the study.

Note that full information on the approval of the study protocol must also be provided in the manuscript.



# Pharmacologic fibroblast reprogramming into photoreceptors restores vision

<https://doi.org/10.1038/s41586-020-2201-4>

Received: 23 October 2018

Accepted: 10 February 2020

Published online: 15 April 2020

 Check for updates

Biraj Mahato<sup>1</sup>, Koray Dogan Kaya<sup>2,8</sup>, Yan Fan<sup>1,8</sup>, Nathalie Sumien<sup>3</sup>, Ritu A. Shetty<sup>3</sup>, Wei Zhang<sup>1</sup>, Delaney Davis<sup>3</sup>, Thomas Mock<sup>3</sup>, Subrata Batabyal<sup>4</sup>, Aiguo Ni<sup>1</sup>, Samarendra Mohanty<sup>4</sup>, Zongchao Han<sup>5</sup>, Rafal Farjo<sup>6</sup>, Michael J. Forster<sup>3</sup>, Anand Swaroop<sup>2</sup> & Sai H. Chavala<sup>1,7</sup>✉

Photoreceptor loss is the final common endpoint in most retinopathies that lead to irreversible blindness, and there are no effective treatments to restore vision<sup>1,2</sup>. Chemical reprogramming of fibroblasts offers an opportunity to reverse vision loss; however, the generation of sensory neuronal subtypes such as photoreceptors remains a challenge. Here we report that the administration of a set of five small molecules can chemically induce the transformation of fibroblasts into rod photoreceptor-like cells. The transplantation of these chemically induced photoreceptor-like cells (CiPCs) into the subretinal space of rod degeneration mice (homozygous for *rd1*, also known as *Pde6b*) leads to partial restoration of the pupil reflex and visual function. We show that mitonuclear communication is a key determining factor for the reprogramming of fibroblasts into CiPCs. Specifically, treatment with these five compounds leads to the translocation of AXIN2 to the mitochondria, which results in the production of reactive oxygen species, the activation of NF- $\kappa$ B and the upregulation of *Ascl1*. We anticipate that CiPCs could have therapeutic potential for restoring vision.

Many retinopathies—such as age-related macular degeneration, diabetic retinopathy and retinitis pigmentosa—ultimately result in the loss of retinal neurons, which leads to irreversible vision loss<sup>1,2</sup>. Stem-cell therapy, using embryonic stem cells or induced pluripotent stem cells, is a promising strategy to replace lost retinal cells and improve vision<sup>3,4</sup>. However, protocols for the derivation of candidate replacement cells are cumbersome and time consuming, presenting a challenge for their use in clinical therapy<sup>5–8</sup>.

Direct reprogramming—using ectopic transcription factors and chemicals—bypasses the requirement for pluripotent cells and has resulted in the generation of neurons, astrocytes and cardiomyocytes; however, the pharmacological conversion of photoreceptors using this method has not been realized<sup>9–13</sup>. ASCL1, a powerful proneural transcription factor, has been reported to reprogram glial cells into photoreceptors<sup>14–16</sup>. An improved mechanistic understanding of direct reprogramming may lead to the generation of new cell types.

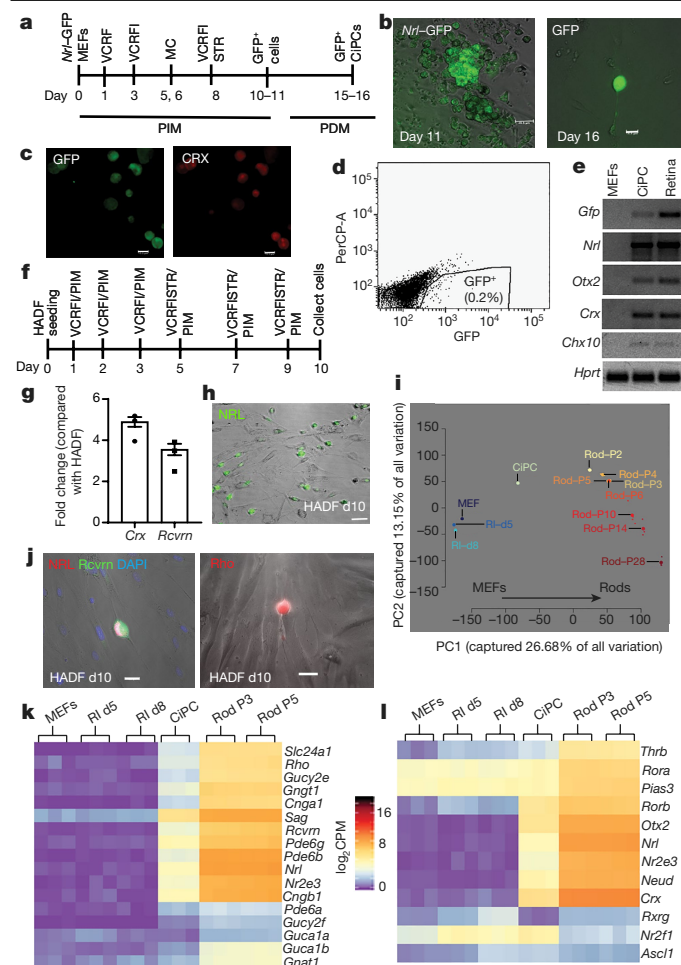
Here we identify a set of five small molecules that can induce fibroblasts to become functional CiPCs without the need for pluripotent cells or viral transcription factors. We demonstrate that CiPCs restore pupil reflex and vision when transplanted into the subretinal space of *rd1* mice, a mouse model of retinal degeneration. Moreover, our mechanistic analysis reveals an AXIN2–NF- $\kappa$ B–ASCL1 pathway that promotes retinal lineage during reprogramming and identifies mitochondria as a signalling hub in the orchestration of cell fate conversion.

## A set of five compounds transforms fibroblasts to CiPCs

To generate CiPCs, we used mouse embryonic fibroblasts (MEFs) derived from a transgenic *Nrl*–GFP mouse, in which the *Nrl* promoter drives expression of eGFP specifically in rod photoreceptors<sup>17,18</sup>. We began by testing an established combination of four small molecules—valproic acid (V), CHIR99021 (a GSK3 inhibitor) (C), RepSox (R) and forskolin (F), together denoted VCRF—that is known to convert fibroblasts into neurons, but only a few of the resultant cells were *Nrl*–GFP positive<sup>12</sup>. Various small molecules and culture conditions were attempted (Supplementary Table 5), and we found that the Wnt/ $\beta$ -catenin pathway inhibitor IWR1 (I)—in combination with VCRF and STR (Sonic hedgehog, taurine and retinoic acid)—was able to substantially improve the efficiency of conversion of MEFs into *Nrl*–GFP<sup>+</sup> cells (Fig. 1a, b). STR was added on day 8 of reprogramming, to promote and support the formation of photoreceptors after photoreceptor-specifying transcription factors—such as ROR $\beta$ , ASCL1, PIAS3—were upregulated<sup>14,19–22</sup> (Fig. 1l, Supplementary Tables 1, 2). Similarly, human adult dermal fibroblasts (HADF), and human fetal lung fibroblasts transduced with a *Nrl*–DsRed reporter, were also converted into CiPCs (Fig. 1f–h, j, Extended Data Fig. 1a, b, Supplementary Fig. 2, Supplementary Table 6). We next tested the small molecules individually and in various combinations, and observed that they failed to generate as many *Nrl*–GFP<sup>+</sup> cells (Extended Data Fig. 1c, d). We conclude that all five compounds in combination can efficiently convert fibroblasts into *NRL*-expressing CiPCs.

<sup>1</sup>Department of Pharmacology and Neuroscience, Laboratory for Retinal Rehabilitation, North Texas Eye Research Institute, University of North Texas Health Science Center, Fort Worth, TX, USA. <sup>2</sup>Neurobiology, Neurodegeneration and Repair Laboratory, National Eye Institute, National Institutes of Health, Bethesda, MD, USA. <sup>3</sup>Department of Pharmacology and Neuroscience, University of North Texas Health Science Center, Fort Worth, TX, USA. <sup>4</sup>Nanoscope Technologies LLC, Bedford, TX, USA. <sup>5</sup>Department of Ophthalmology, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. <sup>6</sup>EyeCRO LLC, Oklahoma City, OK, USA. <sup>7</sup>CIRC Therapeutics, Inc., Dallas, TX, USA. <sup>8</sup>These authors contributed equally: Koray Dogan Kaya, Yan Fan.

✉e-mail: schavala@gmail.com



**Fig. 1 | Conversion of fibroblasts and the molecular characterization of CiPCs.** **a**, Protocol for the reprogramming of mouse fibroblasts into CiPCs. MC, medium change; PIM, photoreceptor induction medium; PDM, photoreceptor differentiation medium. **b**, Images of CiPCs expressing *Nrl*-GFP on day 11 and day 16. Scale bars 14.4  $\mu$ m. **c**, Images of CiPCs expressing CRX on day 11. Scale bars 14.4  $\mu$ m. **d**, FACS purification of reprogrammed *Nrl*-GFP<sup>+</sup> CiPCs (0.2%). **e**, PCR with reverse transcription reveals the expression of the indicated photoreceptor-specific genes in mouse. For gel source data, see Supplementary Fig. 2a, b. **f**, Protocol for the reprogramming of HADF into CiPCs. **g**, Quantitative PCR (qPCR) analysis (fold change compared with HADF) CiPCs after reprogramming from HADF, showing increased expression of photoreceptor-specific genes. Data are presented as mean  $\pm$  s.e.m. of  $n = 3$  independently treated biological replicates. **h**, Micrograph of NRL-stained CiPCs after conversion from HADF. Scale bars 33  $\mu$ m. **i**, Principal component analysis for all RNA-seq samples ( $n = 3$  samples for each). **j**, Images of CiPCs after conversion from HADF at day 10. Rcvrn, recoverin; Rho, rhodopsin. Scale bars 14.4  $\mu$ m. **k**, Heat map of RNA-seq data for the indicated photoreceptor genes. RI, reprogramming intermediate. **l**, Heat map of RNA-seq data for the expression of the indicated genes that express retinal transcription factors during CiPC conversion. Experiments in **b**, **h**, **i** were repeated three times with similar results and the experiments in **c**, **d** were repeated twice with similar results. CPM, counts per million.

Next, *Nrl*-GFP<sup>+</sup> CiPCs were purified by fluorescence-activated cell sorting (FACS) and subject to transcriptomic analysis, which revealed the expression of early retinal neuronal markers (CHX10 and OTX2) and photoreceptor markers (CRX and NRL) (Fig. 1d, e, Supplementary Table 3). Immunostaining analysis of CiPCs on day 11 enabled the visualization of CRX and CHX10 (Fig. 1c, Extended Data Fig. 1c, Supplementary Table 4). Molecular analysis of reprogrammed human CiPCs revealed the expression of photoreceptor-specific markers such

as CRX, rhodopsin and recoverin (Fig. 1g–i, Extended Data Fig. 1c, d, Supplementary Fig. 2).

## CiPCs express photoreceptor genes

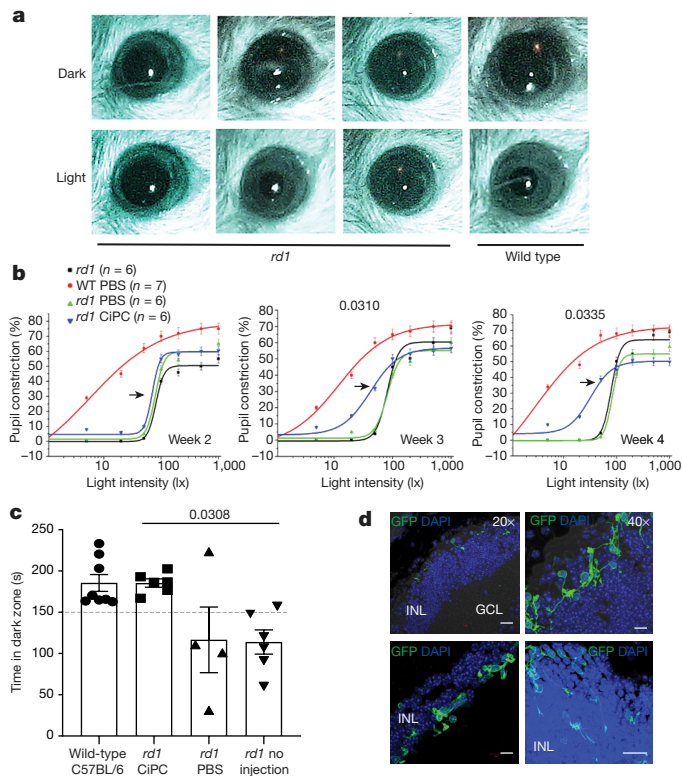
Transcriptome profiling demonstrated that the transcriptome of *Nrl*-GFP<sup>+</sup> CiPCs resembles that of native rods, which were used as a positive control<sup>23</sup>. Heat map analysis revealed the expression of rod-specific genes in CiPCs (Fig. 1i–l, Extended Data Fig. 9a, b), although progenitor-specific genes of other retinal cells—such as cone cells, ganglion cells and Müller glial cells—were expressed at very low levels (Extended Data Fig. 2b). Moreover, fibroblast-specific genes were found to be downregulated in CiPCs and in reprogramming intermediates (Extended Data Fig. 9c). Expression of retinal transcription factors—such as *Otx2*, *Nrl* and *Crx*—was also evident in CiPCs (Fig. 1k). Notably, photoreceptor-specific transcription factors—such as *Rorb*, *Ascl1*, *Pias3*, *Thrb* and *Rxrg*—were upregulated in reprogramming intermediates<sup>21,24,25</sup> (Fig. 1l). These results indicate that CiPCs have a similar gene expression profile to that of their native counterparts.

Because MEFs are composed of a heterogeneous cell population, we performed lineage tracing with *Fsp1*-Cre;Ai9 MEFs (after FACS purification) and confirmed that CiPCs can originate from fibroblasts (Extended Data Fig. 1e–h). Staining with 5-bromodeoxyuridine (BrdU) during reprogramming to CiPCs indicated that approximately 95% of *Nrl*-GFP<sup>+</sup> cells were negative for BrdU, suggesting that no intermediate proliferative stage exists (Extended Data Fig. 2a).

## CiPC transplantation in *rd1* mice

To test whether CiPCs can activate existing retinal circuitry and restore visual function, FACS-purified *Nrl*-GFP<sup>+</sup> CiPCs were transplanted into the subretinal space of 14 *rd1* eyes, a mouse model of retinal degeneration (Extended Data Fig. 2c). Recently, pupillary light reflex has been reported as a robust method to measure photoreceptor function after cell transplantation<sup>26</sup>. Pupillary constriction under low light levels is critically dependent upon the function of rod photoreceptors. Six out of 14 (43%) *rd1* eyes demonstrated improved pupil response under low-light conditions three and four weeks after transplantation (Fig. 2a, b). None of the mice demonstrated pupil constriction at two weeks, which served as a baseline (internal negative control) for longitudinal comparison and reduced the likelihood that a pre-existing photoreceptor or alternative pupillary pathway—such as the melanopsin pathway—is responsible for the observed pupillary restoration (Fig. 2b). To assess the restoration of visual function, the six pupil-responsive mice underwent light-aversion testing. The light-aversion test provides an option for mice to spend time in either a dark space or a lit space. Mice have an innate tendency to avoid lit spaces; as such, those with visual function favour dark spaces<sup>27</sup>. To test rod vision in these pupil-reflex-positive mice, we allowed the mice to adapt to dark conditions and then performed the test using illumination conditions of 50 lx. *rd1* mice injected with CiPCs were found to spend significantly more time in the dark space compared with those injected with PBS control (Fig. 2c). None of the CiPC-injected mice without improved pupil responses demonstrated a dark preference. We also performed a modified optomotor test with neutral density filters; this modification was necessary in order to prevent rod saturation and to test rod-mediated vision. An additional benefit of testing at low light intensities is that the possibility of activating any residual cones is substantially reduced. One of the six CiPC-injected mice that demonstrated improved pupil responses and dark preference also showed improvements in visual acuity and contrast sensitivity (Extended Data Fig. 3e, f). These results provide a proof-of-concept that CiPCs can restore visual function in *rd1* mice.

We next sought to rule out the possibility that CiPCs primarily function to reduce the degeneration of host photoreceptors. We transplanted *Nrl*-GFP<sup>+</sup> CiPCs into the subretinal space of *rd1* mice on day



**Fig. 2 | Functional analysis of CiPCs in a mouse model of retinal degeneration.** **a**, Representative images from the pupil analysis. CiPC-injected eyes 3 months after transplantation. **b**, Illuminance response curves for pupillary constriction ( $n = 6$ ). The stimulus was a 20-s light exposure and the data are fitted with a sigmoidal function. Arrows show the recovery of pupil function. The results presented are for pupil-responsive mice only. Statistical significance was assessed using a two-tailed Student's *t*-test. WT, wild type. **c**, The time spent by mice in the dark area of the light-aversion test. Wild-type C57BL/6 mice were used for comparison. Statistical significance was assessed using a one-way ANOVA. Data are presented as mean  $\pm$  s.e.m. **d**, Top left, integration and survival of GFP<sup>+</sup> CiPCs in the retina of an *rd1* mouse 3 months after transplantation. Scale bar, 20  $\mu$ m. Top right, a magnified view of the top left image. Scale bar, 10  $\mu$ m. Bottom, additional images of the integration of GFP<sup>+</sup> CiPCs. Bottom left, 40 $\times$  magnification; bottom right, z-section. Scale bars: left, 10  $\mu$ m; right, 20  $\mu$ m. GCL, ganglion cell layer; INL, inner nuclear layer. Experiments in **a** and **d** were repeated twice with similar results and *n* denotes the number of biologically independent mice.

31—a later time point at which there is near complete photoreceptor degeneration and electroretinogram (ERG) signals are extinguished<sup>28</sup>. Long-term retinal function was analysed by pupillary light reflex and ERG (Extended Data Fig. 2c). Approximately three months after transplantation (at postnatal day (P)128), we recorded a 30%–40% increase in pupillary constriction in response to a low light stimulus (50 lx) in three of the six eyes into which CiPCs had been transplanted (Extended Data Figs. 2c and 3d). Full-field ERG analysis demonstrated an improvement of the scotopic a-wave in three out of six eyes into which CiPCs had been transplanted at P45; this improvement was lost at and after P59 (Extended Data Fig. 3a–c). One eye showed improvement in the scotopic b-wave at P45 (Extended Data Fig. 2d).

The long-term survival of transplanted CiPCs and their synaptic connections to inner retinal neurons were assessed by immunofluorescence. Three months after CiPC transplantation, we identified GFP<sup>+</sup>, GFP<sup>+</sup>recoverin<sup>+</sup> or GFP<sup>+</sup>rhodopsin<sup>+</sup> CiPCs in the outer aspect of the inner nuclear layer in eyes that demonstrated restoration of the pupillary light reflex (Fig. 2d, Extended Data Figs. 4a, b, 8d). We quantified CiPC survival and found a strong correlation between pupil constriction and

cell survival (Extended Data Fig. 8e). Specifically, eyes with improved pupil response and light/dark response had an average of 58 CiPCs per section, compared with 8 CiPCs per section in non-responders (Extended Data Fig. 8f). Moreover, CiPCs were found to have synaptic terminals that express the rod ribbon-synapse protein ribeye and the synaptic vesicle protein synaptophysin. These synaptic terminals are in close proximity to rod bipolar cells (expressing bipolar marker PKC- $\alpha$ ), which is essential for transmitting a light signal into the inner retina (Extended Data Fig. 4c–e). Taken together, these results suggest that some of the transplanted CiPCs survive, function and connect with the inner retinal neurons of *rd1* mice.

## NF- $\kappa$ B induces ASCL1 during conversion to CiPCs

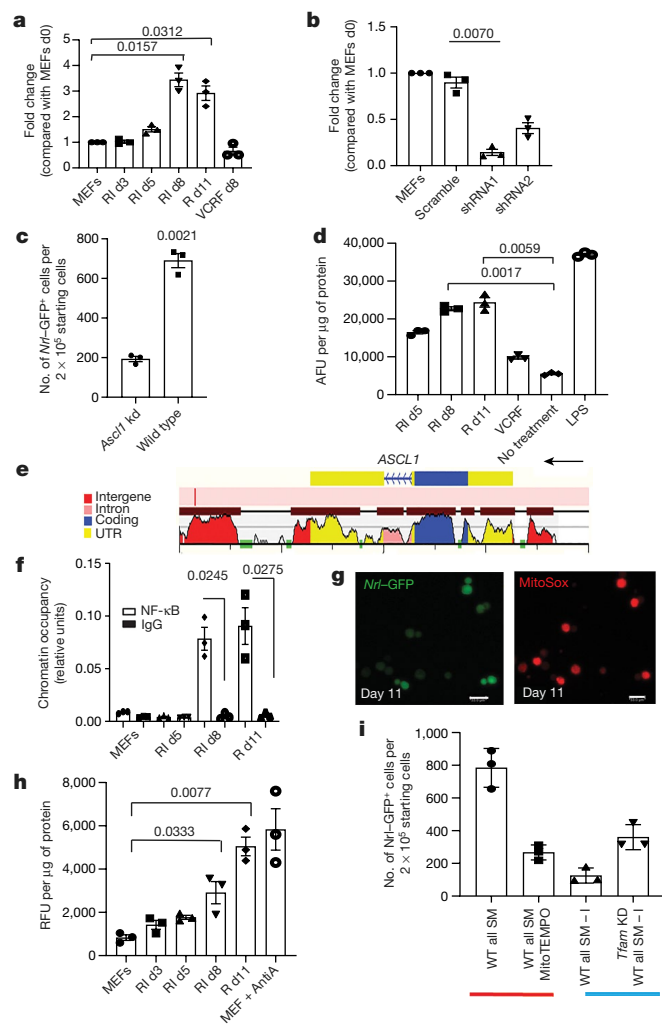
To explore the mechanism of reprogramming to CiPCs, we identified candidate transcription factors from RNA sequencing (RNA-seq) analysis. We detected the expression of *Ascl1* on day 5, which continued to increase until day 8 (Fig. 1l, Fig. 3a). ASCL1, a proneural transcription factor, is capable of converting fibroblasts into neurons<sup>11,29</sup>. Furthermore, ASCL1 is reported to be transiently expressed in photoreceptor precursors and can reprogram Müller glia into rod photoreceptor-like cells<sup>14,15,30</sup>. Taken together, we proposed that the *Ascl1* transcriptional network has a central role in reprogramming fibroblasts to CiPCs. To test this hypothesis, we reprogrammed *Ascl1*-depleted (Fig. 3b) *Nrl*-GFP MEFs with the set of five compounds (VCRFI). We noted a 70%–80% reduction in the generation of CiPCs, which suggests that ASCL1 has an important role in the reprogramming of fibroblasts to CiPCs (Fig. 3c).

To determine the mechanism of ASCL1 induction, we investigated potential upstream regulators of ASCL1. NF- $\kappa$ B is a rapidly acting primary transcription factor that is known to have a role in neural stem-cell differentiation and embryonic neurogenesis<sup>31</sup>. We therefore postulated that NF- $\kappa$ B is an upstream regulator of ASCL1 induction. A luciferase assay demonstrated that NF- $\kappa$ B activation began on day 5 and reached a maximum at day 11 (Fig. 3d). These results suggested the involvement of NF- $\kappa$ B, and so we next explored whether NF- $\kappa$ B induces the expression of ASCL1. Bioinformatics analysis (using rVista) identified a putative binding site for NF- $\kappa$ B near the *Ascl1* loci (Fig. 3e), and a chromatin immunoprecipitation (ChIP) assay confirmed the binding of NF- $\kappa$ B at this locus (Fig. 3f). Furthermore, transient transfection analysis with a luciferase reporter gene confirmed that NF- $\kappa$ B positively regulates *Ascl1* during reprogramming to CiPCs. NF- $\kappa$ B depletion in MEFs that were subsequently reprogrammed to CiPCs resulted in reduced expression of *Ascl1* and fewer CiPCs overall, which further confirms the pathway (Extended Data Fig. 5a–c). Overexpression of *Ascl1* alone in MEFs was not sufficient for reprogramming to CiPCs (Extended Data Fig. 5f–h). These results indicate that treatment with our set of five compounds leads to the activation of NF- $\kappa$ B, which in turn binds to the regulatory regions of *Ascl1* and controls its expression. Notably, analysis of CiPCs using an assay for transposase-accessible chromatin with sequencing (ATAC-seq) revealed the presence of open chromatin at the upstream regions of the *Ascl1* gene in reprogramming intermediates. Homer analysis revealed the enrichment of *Ascl1* and NF- $\kappa$ B-binding motifs in both intermediates and reprogrammed cells, which may be important for the binding of regulatory transcription factors and for their expression (Extended Data Fig. 9f, Supplementary Table 7).

## Reactive oxygen species activate NF- $\kappa$ B

To investigate the mechanism by which NF- $\kappa$ B activation is induced by the combination of five small molecules, we considered known inducers of NF- $\kappa$ B—such as TNF $\alpha$ , lipopolysaccharide, ionizing radiation and mitochondrial reactive oxygen species (mROS)—as possible candidates<sup>32,33</sup>. mROS have been reported to induce nuclear gene expression through the activation of NF- $\kappa$ B<sup>32</sup>. We therefore postulated that mROS generated by treatment with the five compounds may activate NF- $\kappa$ B.





**Fig. 3 | The mROS–NF-κB–ASCL1 signalling axis determines the reprogramming of fibroblasts to CiPCs.** **a**, *Ascl1* transcript expression during reprogramming of MEFs to CiPCs, analysed by qPCR using the  $2^{-\Delta\Delta CT}$  method. RI, reprogramming intermediates, R, reprogrammed cells. **b**, *Ascl1* qPCR analysis in *Ascl1* depleted MEFs, using short hairpin RNA (shRNA). **c**, Number of GFP<sup>+</sup> cells after reprogramming to CiPCs from *Ascl1*-knockdown (KD) and wild-type *Nrl*-GFP MEFs. **d**, NF-κB-luciferase activity during reprogramming of MEFs to CiPCs. LPS, lipopolysaccharide. **e**, rVista sequence alignment of human *ASCL1* and mouse *Ascl1* genes show highly conserved NF-κB-binding sites (red vertical bar) downstream of the 3'UTR region. **f**, ChIP assay shows binding of NF-κB at downstream loci of *Ascl1*. **g**, Accumulation of mROS in CiPCs converted from *Nrl*-GFP MEFs on day 11. Scale bar 33 μm. **h**, Fluorometric analysis of mROS generation during reprogramming of MEFs to CiPCs. AntiA, antimycin A; RFU, relative fluorescence units. **i**, Quantification of *Nrl*-GFP<sup>+</sup> cells on day 11 after depletion (by MitoTEMPO treatment) or generation (by *Tfam* knockdown) of mROS. SM, small molecules. In **a–d**, **f**, **h**, **i**, data are presented as mean ± s.e.m. of  $n = 3$  independently treated biological replicates. Statistical significance was assessed by two-tailed Student's *t*-test.

We observed an evident increase in mROS accumulation, beginning on day 8 in reprogramming intermediates (Fig. 3g, h). To determine the importance of mROS, we used the antioxidant MitoTEMPO, an mROS scavenger, and observed a reduction in reprogramming to CiPCs (Fig. 3i). Depletion of mitochondrial transcription factor A (TFAM) has been reported to lead to the generation of mROS<sup>34</sup>. Our data show that withdrawal of the compound IWR1 reduces reprogramming efficiency and mROS generation (Fig. 3i, Extended Data Fig. 6f). We considered the possibility that TFAM-depleted MEFs may have increased reprogramming potential in the absence of IWR1. Consistent with this hypothesis,

we found increased reprogramming to CiPCs and increased mROS generation in TFAM-depleted MEFs in the absence of IWR1 (Fig. 3i, Extended Data Fig. 6e, f). Furthermore, exogenous ROS or TNFα (a NF-κB inducer) did not have a significant effect on CiPC conversion efficiency, and excessive ROS had a negative effect (Extended Data Fig. 5d, e). These data indicate that mROS has an important role in the reprogramming of fibroblasts to CiPCs.

To determine whether the activation of NF-κB is dependent on mROS, we added MitoTEMPO on day 3 of reprogramming and observed significantly decreased activity in the luciferase assay (Extended Data Fig. 6a). Additionally, we found that the extent of binding of NF-κB near the *Ascl1* loci was reduced upon treatment with MitoTEMPO (Extended Data Fig. 6g). Taken together, these results demonstrate that mitochondrial ROS activates NF-κB, which controls the expression of *Ascl1* by binding its regulatory region.

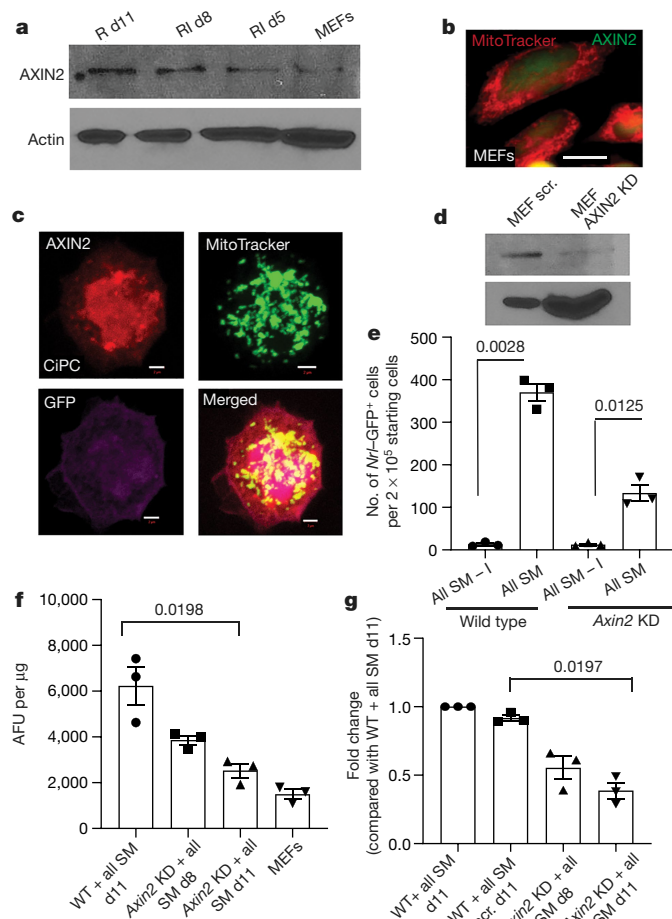
### The five compounds promote Axin2 mitolocalization

To identify the mechanism by which mROS is generated during reprogramming to CiPCs, we considered two reports that demonstrate the stabilization of AXIN2 by treatment with CHIR and IWR1<sup>35</sup> and the translocation of AXIN2 to mitochondria upon treatment with XAV939 (an IWR1 analogue)<sup>36</sup>. We proposed that treatment with the group of five compounds—which includes IWR1 and CHIR—induces the stabilization of AXIN2 and its subsequent translocation to mitochondria; mitochondria-targeted AXIN2 then generates mROS, which in turn activates NF-κB. To test the hypothesis, the expression of AXIN2 was examined during the reprogramming to CiPCs. We discovered that AXIN2 in reprogramming intermediates and in day 11 non-purified CiPCs is more stabilized compared with that in starting MEFs (Fig. 4a). Subsequently, we found that stabilized AXIN2 translocates to the mitochondria of CiPCs, as evidenced by its co-localization with mitochondria (Fig. 4b, c, Extended Data Fig. 7a, b). *Axin2*-depleted MEFs demonstrate reduced reprogramming to CiPCs and reduced mROS generation (Fig. 4d–f, Extended Data Fig. 6d). For further confirmation we measured NF-κB activation and *Ascl1* expression in *Axin2*-depleted reprogramming intermediates. We detected reduced *Ascl1* expression and decreased NF-κB activity (Fig. 4g, Extended Data Fig. 6b). Taken together, these results indicate that mitochondria-targeted AXIN2 induces increased mROS generation in CiPCs and activates the downstream NF-κB–ASCL1 pathway to promote a lineage switch to a photoreceptor fate. Moreover, mitochondrial analysis of CiPCs revealed low basal mitochondrial respiration, low ATP turnover, low reserve capacity and high glycolysis rates; this is indicative of an immature mitochondrial state, which may be linked to the generation of mROS (Extended Data Fig. 7c–f).

### Discussion

Here we report that a combination of five small molecules can convert fibroblasts into functional CiPCs that are capable of partially restoring pupil reflex and visual function in a mouse model of retinal degeneration (Extended Data Fig. 10a). Gene expression profiling reveals that CiPCs are similar to their in vivo rod counterpart, and CiPC conversion recapitulates the ontogeny of photoreceptor genesis—as indicated by the upregulation of *Thrb*, *Rorb* and *Pias3*, among others<sup>21,24</sup>. The occurrence of open chromatin regions near photoreceptor loci and the enrichment of photoreceptor-specifying transcription-factor-binding motifs during reprogramming to CiPCs further validate these results<sup>37,38</sup> (Extended Data Fig. 9f, Supplementary Table 7). Subretinal transplantation of rod-like CiPCs into *rd1* mice led to long-term improvement in pupillary light reflex and the restoration of normal visual behaviour in the light aversion test. In our study, 6 out of 14 mice (43%) demonstrated an improved pupillary response. The strong correlation between pupil constriction and cell survival may explain why some eyes showed an improvement and others did not (Extended Data Fig. 8e).





**Fig. 4 | Mitochondria-translocated AXIN2 causes mROS generation and the reprogramming of fibroblasts to CiPCs.** **a**, Western blot demonstrating AXIN2 expression during reprogramming of MEFs to CiPCs. For gel source data, see Supplementary Fig. 1c. **b**, No co-localization of AXIN2 (green) with mitochondria (MitoTracker) was observed in MEFs. Scale bar, 13.4  $\mu$ m. **c**, AXIN2 (red) localization in mitochondria (green) of GFP<sup>+</sup> CiPCs (purple pseudocolour). Scale bar, 2  $\mu$ m. **d**, Western blot demonstrating reduced levels of AXIN2 after knockdown of its encoding gene, *Axin2*, by shRNA. For gel source data, see Supplementary Fig. 1g. **e**, Quantification of *Nrl*-GFP<sup>+</sup> cells after conversion of *Axin2*-depleted MEFs to CiPCs on day 11. **f**, Reduced mROS generation in *Axin2*-depleted cells during reprogramming to CiPCs, as assessed by fluorimetry. AFU, arbitrary fluorescence units. **g**, qPCR analysis shows that knockdown of *Axin2* is associated with reduced expression of *Ascl1* in reprogramming intermediates on day 8 and CiPCs on day 11. Scr., scramble. In **e–g**, data are presented as mean  $\pm$  s.e.m. of  $n = 3$  independently treated biological replicates. Statistical significance was assessed by two-tailed Student's *t*-test.

Material transfer and cell fusion are mechanisms that can explain improvement in visual function; however, these mechanisms have not been reported in the late stages of retinal degeneration, during which viable host photoreceptors are required in order for improvement to occur<sup>26,39</sup>. These mechanisms are of most concern in photoreceptor loss-of-function models—such as *Gnat1*—in which the photoreceptors have limited or no degeneration and retinal morphology and structure are preserved. To avoid the possibility of donor–host cell fusion and material transfer, we chose to transplant CiPCs into *rd1* mice at a time point (postnatal day (P)31) at which no rod photoreceptors exist<sup>28,40</sup>. Although rod loss is almost complete at P31, we examined the possibility that existing cones or bipolar cells could participate in CiPC material

transfer or cell fusion; however, we found that these are unlikely to have a major role (Extended Data Figs. 4d, 8c).

Our mechanistic studies reveal that mROS-mediated NF- $\kappa$ B activation directly regulates the expression of *Ascl1* and the reprogramming of fibroblasts into CiPCs. Mitochondrial signalling is reported to have a role in cellular homeostasis and neuronal function<sup>41–43</sup>. To our knowledge, this report is the first to demonstrate that mitochondria-to-nucleus signalling acts as a mediator for direct chemical reprogramming (Extended Data Fig. 10b). Moreover, mitochondria-translocated AXIN2 probably generates mROS, although the mechanism is currently unknown. Induction of the reprogramming of fibroblasts to CiPCs by our group of five compounds reveals a new function of the mitochondria that may provide valuable knowledge for the generation of other cell types.

Although CiPCs have therapeutic potential, a lack of proliferation—as is the case for native photoreceptors—and low conversion efficiency are the main impediments for a translational application. We anticipate that optimization of our current protocol may be beneficial for obtaining large numbers of CiPCs. For example, temporal modulation of IWR1 in the protocol resulted in a significant increase in the conversion of HADF to CiPCs (Fig. 1f, Extended Data Fig. 8d, right). Overall, CiPCs are a promising cell-replacement candidate and may lead to a scalable therapy for vision restoration.

## Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-020-2201-4>.

- Wright, A. F., Chakarova, C. F., Abd El-Aziz, M. M. & Bhattacharya, S. S. Photoreceptor degeneration: genetic and mechanistic dissection of a complex trait. *Nat. Rev. Genet.* **11**, 273–284 (2010).
- Bramall, A. N., Wright, A. F., Jacobson, S. G. & McInnes, R. R. The genomic, biochemical, and cellular responses of the retina in inherited photoreceptor degenerations and prospects for the treatment of these disorders. *Annu. Rev. Neurosci.* **33**, 441–472 (2010).
- Schwartz, S. D. et al. Human embryonic-stem cell derived retinal pigment epithelium in patients with age-related macular degeneration and Stargardt's macular dystrophy: follow up of two open-label phase 1/2 studies. *Lancet* **385**, 509–516 (2015).
- Dyer, M. A. An eye on retinal recovery. *Nature* **540**, 350–351 (2016).
- Mellough, C. B., Sernagor, E., Moreno-Gimeno, I., Steel, D. H. & Lako, M. Efficient stage-specific differentiation of human pluripotent stem cells toward retinal photoreceptor cells. *Stem Cells* **30**, 673–686 (2012).
- Zhong, X. et al. Generation of three-dimensional retinal tissue with functional photoreceptors from human iPSCs. *Nat. Commun.* **5**, 4047 (2014).
- Gamm, D. M. & Wright, L. S. From embryonic stem cells to mature photoreceptors. *Nat. Biotechnol.* **31**, 712–713 (2013).
- Völkner, M. et al. Retinal organoids from pluripotent stem cells efficiently recapitulate retinogenesis. *Stem Cell Reports* **6**, 525–538 (2016).
- Fu, Y. et al. Direct reprogramming of mouse fibroblasts into cardiomyocytes with chemical cocktails. *Cell Res.* **25**, 1013–1024 (2015).
- Tian, E. et al. Small-molecule-based lineage reprogramming creates functional astrocytes. *Cell Rep.* **16**, 781–792 (2016).
- Vierbuchen, T. et al. Direct conversion of fibroblasts to functional neurons by defined factors. *Nature* **463**, 1035–1041 (2010).
- Babos, K. & Ichida, J. K. Small molecules take a big step by converting fibroblasts into neurons. *Cell Stem Cell* **17**, 127–129 (2015).
- Zhang, L. et al. Small molecules efficiently reprogram human astroglial cells into functional neurons. *Cell Stem Cell* **17**, 735–747 (2015).
- Ueki, Y. et al. Transgenic expression of the proneural transcription factor *Ascl1* in Müller glia stimulates retinal regeneration in young mice. *Proc. Natl Acad. Sci. USA* **112**, 13717–13722 (2015).
- Jorstad, N. L. et al. Stimulation of functional neuronal regeneration from Müller glia in adult mice. *Nature* **548**, 103–107 (2017).
- Pollak, J. et al. ASCL1 reprograms mouse Müller glia into neurogenic retinal progenitors. *Development* **140**, 2619–2631 (2013).
- Akimoto, M. et al. Targeting of GFP to newborn rods by *Nrl* promoter and temporal expression profiling of flow-sorted photoreceptors. *Proc. Natl Acad. Sci. USA* **103**, 3890–3895 (2006).
- Kim, J. W. et al. NRL-regulated transcriptome dynamics of developing rod photoreceptors. *Cell Rep.* **17**, 2460–2473 (2016).
- Osakada, F. et al. Toward the generation of rod and cone photoreceptors from mouse, monkey and human embryonic stem cells. *Nat. Biotechnol.* **26**, 215–224 (2008).

20. Khanna, H. et al. Retinoic acid regulates the expression of photoreceptor transcription factor NRL. *J. Biol. Chem.* **281**, 27327–27334 (2006).
21. Onishi, A. et al. Pias3-dependent SUMOylation directs rod photoreceptor development. *Neuron* **61**, 234–246 (2009).
22. Fu, Y. et al. Feedback induction of a photoreceptor-specific isoform of retinoid-related orphan nuclear receptor  $\beta$  by the rod transcription factor NRL. *J. Biol. Chem.* **289**, 32469–32480 (2014).
23. Kim, J. W. et al. Recruitment of rod photoreceptors from short-wavelength-sensitive cones during the evolution of nocturnal vision in mammals. *Dev. Cell* **37**, 520–532 (2016).
24. Jia, L. et al. Retinoid-related orphan nuclear receptor ROR is an early-acting factor in rod photoreceptor development. *Proc. Natl Acad. Sci. USA* **106**, 17534–17539 (2009).
25. Brzezinski, J. A., IV, Kim, E. J., Johnson, J. E. & Reh, T. A. Ascl1 expression defines a subpopulation of lineage-restricted progenitors in the mammalian retina. *Development* **138**, 3519–3531 (2011).
26. Zhu, J., Cifuentes, H., Reynolds, J. & Lamba, D. A. Immunosuppression via loss of IL2ry enhances long-term functional integration of hESC-derived photoreceptors in the mouse retina. *Cell Stem Cell* **20**, 374–384.e5 (2017).
27. Lin, B., Koizumi, A., Tanaka, N., Panda, S. & Masland, R. H. Restoration of visual function in retinal degeneration mice by ectopic expression of melanopsin. *Proc. Natl Acad. Sci. USA* **105**, 16009–16014 (2008).
28. Nishiguchi, K. M. et al. Gene therapy restores vision in *rd1* mice after removal of a confounding mutation in *Gpr179*. *Nat. Commun.* **6**, 6006 (2015).
29. Pang, Z. P. et al. Induction of human neuronal cells by defined transcription factors. *Nature* **476**, 220–223 (2011).
30. Swaroop, A., Kim, D. & Forrest, D. Transcriptional regulation of photoreceptor development and homeostasis in the mammalian retina. *Nat. Rev. Neurosci.* **11**, 563–576 (2010).
31. Zhang, Y. et al. Nuclear factor kappa B signaling initiates early differentiation of neural stem cells. *Stem Cells* **30**, 510–524 (2012).
32. Formentini, L., Sánchez-Aragó, M., Sánchez-Cenizo, L. & Cuezva, J. M. The mitochondrial ATPase inhibitory factor 1 triggers a ROS-mediated retrograde prosurvival and proliferative response. *Mol. Cell* **45**, 731–742 (2012).
33. Andreaskos, E. et al. Distinct pathways of LPS-induced NF-kappa B activation and cytokine production in human myeloid and nonmyeloid cells defined by selective utilization of MyD88 and Mal/TIRAP. *Blood* **103**, 2229–2237 (2004).
34. Vernochet, C. et al. Adipose-specific deletion of TFAM increases mitochondrial oxidation and protects mice against obesity and insulin resistance. *Cell Metab.* **16**, 765–776 (2012).
35. Kim, H. et al. Modulation of  $\beta$ -catenin function maintains mouse epiblast stem cell and human embryonic stem cell self-renewal. *Nat. Commun.* **4**, 2403 (2013).
36. Shin, J. H., Kim, H. W., Rhyu, I. J. & Kee, S. H. Axin is expressed in mitochondria and suppresses mitochondrial ATP synthesis in HeLa cells. *Exp. Cell Res.* **340**, 12–21 (2016).
37. Rehemtulla, A. et al. The basic motif-leucine zipper transcription factor Nrl can positively regulate rhodopsin gene expression. *Proc. Natl Acad. Sci. USA* **93**, 191–195 (1996).
38. Morrow, E. M., Furukawa, T., Lee, J. E. & Cepko, C. L. NeuroD regulates multiple functions in the developing neural retina in rodent. *Development* **126**, 23–36 (1999).
39. Ortin-Martinez, A. et al. A reinterpretation of cell transplantation: GFP transfer from donor to host photoreceptors. *Stem Cells* **35**, 932–939 (2017).
40. Singh, M. S. et al. Reversal of end-stage retinal degeneration and restoration of visual function by photoreceptor transplantation. *Proc. Natl Acad. Sci. USA* **110**, 1101–1106 (2013).
41. Zhou, G., Meng, S., Li, Y., Ghebre, Y. T. & Cooke, J. P. Optimal ROS signaling is critical for nuclear reprogramming. *Cell Rep.* **15**, 919–925 (2016).
42. Shadel, G. S. & Horvath, T. L. Mitochondrial ROS signaling in organismal homeostasis. *Cell* **163**, 560–569 (2015).
43. Cagin, U., Duncan, O. F., Gatt, A. P., Dionne, M. S., Sweeney, S. T. & Bateman, J. M. Mitochondrial retrograde signaling regulates neuronal function. *Proc. Natl Acad. Sci. USA* **112**, E6000–E6009 (2015).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2020

## Methods

### Generation of chemically induced photoreceptor-like cells

All small molecules were diluted in DMSO or DMEM according to the manufacturer's instructions. Approximately 50,000–80,000 MEFs (passage 2) were seeded into each well of a 12-well plate (0.1% gelatin-coated overnight) and cultured overnight. On day 1, the medium was replaced with photoreceptor induction medium (PIM; Supplementary Tables 1, 2) containing V (0.5 mM), C (4.8  $\mu$ M), R (2  $\mu$ M) and F (10  $\mu$ M). On day 3, fresh PIM containing V, C, R, F plus I (10  $\mu$ M) were added into each well. The medium was then replaced with PIM containing VCRF1 on days 4–7 depending on cell appearance. S (3 nM), T (100  $\mu$ M) and R (1  $\mu$ M) were added on day 8 with fresh medium containing all the small molecules mentioned above. On day 10–11, GFP<sup>+</sup> cells were collected and analysed for gene expression. For CiPC maturation, PIM was replaced with photoreceptor differentiation medium with the small molecules VCRF and cultured up to day 15–16 (Supplementary Tables 1, 2). CiPC survival decreases considerably after day 11 with only a few mature cells obtained at day 15–16. For human adult dermal fibroblast (HADF) reprogramming, cells were seeded (to 95 to 100% confluence) in IMR90 medium. On day 1, PIM containing VCRF1 (at the same concentrations as stated above) was added to the wells. PIM containing VCRF1STR was added on day 5. NRL staining was performed between day 7–10. We noted more autofluorescent cells after fixation and these cells were not considered as reprogrammed. The medium and all small molecules were replenished daily throughout the conversion period depending upon appearance. Human Fetal Lung Fibroblasts (HFL1) conversion to CiPCs was similar to mouse protocol, except that 5  $\mu$ M IWR1 was added on day 2. The small molecules S, T and R, as well as brain-derived neurotrophic factor, glial cell line-derived neurotrophic factor and neurotrophin-3 were added on day 5. On day 6–8, DsRed<sup>+</sup> cells were collected for analysis.

### Mouse models and MEF isolation and lineage tracing

All animal studies and animal care were performed in accordance with relevant guidelines and regulations and approved by the Institutional Animal Care and Use Committee at the University of North Texas Health Science Center. *Nrl*-GFP reporter mice were a gift from A. Swaroop (National Eye Institute, NEI) and were used to generate MEFs. *Nrl*-GFP cells were not identified in either starting fibroblast culture<sup>44</sup>. rtTa mice (Jackson Laboratory, 006965) were crossed with *Nrl*-GFP mice for unrelated experiments. Embryos from *Fsp1*-Cre mice (Jackson Laboratory, stock# 012641) crossed with R26-LSL-tdTomato mice (Jackson Laboratory, stock# 007909) were used for lineage tracing. Embryos were examined for tdTomato expression before isolation of MEFs. td-Tomato<sup>+</sup> cells obtained from FACS were used for reprogramming to CiPCs according to the protocol described above, and CiPCs were stained with anti-NRL antibody according to the method detailed below. CBA/J (*rd1*) mice (stock# 000656) and C57BL/6 mice (stock# 000664) were purchased from the Jackson Laboratory.

### Subretinal injection of CiPCs into *rd1* mice

After conversion, GFP<sup>+</sup> CiPCs were sorted by FACS from multiple 10-cm dishes and resuspended in IMR90 medium. For subretinal injection, mice (P31 and P24) were first anaesthetized by intramuscular injection of 85 mg kg<sup>-1</sup> ketamine and 14 mg kg<sup>-1</sup> xylazine. A 30-gauge blunt-end needle attached to a 10 ml Nanofil syringe (World Precision Instruments) was inserted into the puncture site (0.5–1 mm under the limbus line) with visualization aided by the use of a surgical microscope (Leica). CiPCs (80,000; 1.5–2  $\mu$ l per eye) or saline (1.5–2  $\mu$ l per eye) were delivered into the subretinal space in the superior temporal quadrant. After injection, the needle was left in place for a few seconds to reduce reflux and enable maximal cell release before being slowly withdrawn. The eyelid was returned to its original position and a drop of Triple Antibiotic (Equate, Walmart) ointment was applied. Mice were warmed on a 37 °C bed until fully awake.

### Pupillometry

Similar to a previously published protocol<sup>26</sup>, dark-adapted mice were used to capture images under infrared illumination. CiPC-transplanted eyes that showed ERG improvement were exposed to low-irradiance white light through a light guide from a 100-W goose arc lamp at a range of intensities. For the illuminance curve experiments, dark-adapted mice were subjected to a series of light exposures (with increasing illuminance) for 10 s on weeks 2–4 after transplantation. A complete intensity series was performed in one eye before retesting the other eye at similar intensities. A gap of at least 2 min was maintained in between measurements. Pupil constriction was imaged in the contralateral, non-transplanted eyes. An infrared light-emitting diode was used throughout the experiment for background illumination. An infrared camera (Sony, DCR-HC96) was used to acquire images. The pupil area for each eye was measured before and after light exposure with ImageJ software (National Institutes of Health). For each mouse, the change in pupil constriction was represented by the difference between the pupil area measured in the dark and in the light.

### Quantitative PCR and PCR with reverse transcription

Total RNA was extracted using a kit (Zymo Research, Cat# R1050) according to the manufacturer's instructions. RNA (1  $\mu$ g) was converted to cDNA using a High Capacity cDNA Reverse Transcription kit (Applied Biosystems, 4368814). Isolated RNAs were treated with DNAase-I before c-DNA synthesis. A thermal cycler from Applied Biosystems and OneStep Plus real-time PCR were used for amplification. qPCR was performed using Fast SYBR Green Master Mix (Applied Biosystems, 4385612). Results were normalized with glyceraldehyde 3-phosphate dehydrogenase or hypoxanthine-guanine phosphoribosyltransferase and fold change was calculated using the 2<sup>- $\Delta\Delta$ CT</sup> method. A list of primers is provided in Supplementary Table 3.

### Transcriptome analysis

**Quantification.** RNA-seq data analysis was performed at the gene level as previously<sup>45</sup> with Ensembl data release 87. The gene level count matrix was then TMM-normalized (TMM, trimmed mean of M values) using the edgeR (v3.18.1) package in the R (v3.4.0) programming environment, as previously described<sup>46</sup>.

### Gene expression clustering and heat maps

Gene expression clustering was performed on selected genes using the affinity propagation algorithm by using negative distance as the similarity for ordering the genes in each set before drawing the heat maps. The heat map function was developed in-house and is available upon request.

### ATAC-seq

Cells were washed twice with fresh medium and 1:100 volume of 100X DNase buffer (250 mM MgCl<sub>2</sub> and 50 mM CaCl in H<sub>2</sub>O) and 1:100 volume of 100X DNase solution (Worthington, LS002006, resuspended in HBSS at 20,000 units per ml) was added to the medium. Cells were incubated in DNase at 37 °C for 30 min in a tissue culture incubator. After washing the medium with PBS, cells were then trypsinized, centrifuged (800g for 5 min) and resuspended in 500  $\mu$ l of growth medium containing 5% DMSO in a slow cooling chamber. Approximately 100,000 cells were then shipped to Active Motif to perform ATAC-seq and bioinformatics analysis.

### Immunohistochemistry, subcellular fractionation and immunoblotting

Paraformaldehyde (4%) was used to fix enucleated eyes. Cryo-embedded eyes were then sectioned with a thickness of 14  $\mu$ m. Fixed eye sections were analysed with primary and secondary antibodies listed in Supplementary Table 4. DAPI (0.1%) was used to stain the nucleus in mounting

# Article

medium. Images were taken using a Zeiss LSM510 confocal/Leica DMI8 fluorescence microscope. Subcellular fractionation was performed following a kit protocol from Thermo Fisher Scientific (Cat# 89874). For immunoblotting, total protein was extracted with commercially available lysis buffer (Thermo Scientific, Cat# 89900) and then concentration of proteins was measured using a BCA protein assay kit (Thermo Scientific, Cat# 23227). An equivalent amount of proteins was loaded into each well, immunoblotted and antibody-stained using a standard procedure (Supplementary Table 4). SuperSignal West Femto maximum sensitivity substrate (Thermo Scientific, Cat# 34094) was used to develop the subcellular fractionation western blot.

## Electroretinogram

After a minimum dark-adaptation period of 12 h, mice were anaesthetized by intraperitoneal injection of 85 mg kg<sup>-1</sup> ketamine and 14 mg kg<sup>-1</sup> xylazine. Preparation was performed under a dim red light (<50 lux). ERG analyses were performed using an Espion system (Diagnosys). For the assessment of scotopic response, a stimulus intensity of 40 cd·s m<sup>-2</sup> was presented to the dark-adapted dilated eyes. The amplitude of the scotopic a-wave was then measured from the pre-stimulus baseline to the a-wave trough. The amplitude of the b-wave was then measured from the trough of the a-wave to the crest of the b-wave. A total of 15 repeated flashes and measurements were averaged to produce the final waveform. The amplitude of the photopic b-wave was then measured from the trough of the a-wave to the crest of the b-wave. At the beginning of the day, the response of wild-type C57BL/6 mice (aged over P21,  $n \geq 2$ ) was recorded and quantified to ensure proper device calibration.

## Light-aversion test

The visual discrimination (light/dark) test was conducted in an apparatus that consists of black opaque (100%) acrylic test chambers (30.48 × 15.24 × 30.48 cm (length, width, height)). This chamber was further divided into equal-sized compartments (15.24 × 15.24 × 30.48 cm) by the addition of an insert, to create a dividing wall in the centre. Further, to create light and dark zones, one compartment was illuminated with dim ambient light (50 ± 1.5 lx) and the other compartment was kept dark (~0.1 lx). The light and dark compartments were connected by an opening (5 × 5 cm). The position of the mouse within the apparatus was recorded using a photocell-based system (Model 71-CPPX, Omnitech). The acrylic chambers were housed separately in sound-attenuating chambers (Model 71-ECC, Omnitech). Ambient noise within the chambers was 64 dB and testing took place under dim illumination.

Mice were maintained in the testing room overnight (about 12 h) in dark conditions in their home cage with free access to food and water. Each mouse was allowed to habituate to the testing apparatus (both sides) for 10 min while in the dark. After habituation, one side of the apparatus was illuminated with an ambient light at around 50 lx and the mouse was allowed to roam freely between each compartment for 5 min. The time spent in the dark and light compartments was recorded by a photocell-based system.

## Optomotor task

The testing apparatus was a chamber (39 × 39 × 32.5 cm) with mirrored floors and ceilings. Attached to each of the four walls was a 20-inch computer monitor facing inwards. In the centre of the chamber was a platform (7-cm diameter) that was elevated approximately 15 cm from the floor. When a mouse was placed on the platform, a video camera positioned in the ceiling of the apparatus enabled the behaviour of the mouse to be clearly visible during testing. A computer program was used to project visual stimuli (vertical gratings) onto the monitors (OptoMotry, CerebralMechanics). The gratings were rotated at 12° s<sup>-1</sup>, producing the appearance of a virtual rotating cylinder. The moving gratings elicited a tracking behaviour and the visual acuity threshold and contrast sensitivity was determined for each eye rotating in a clockwise direction (when testing the left eye) or in the anticlockwise

direction (when testing the right eye). Light levels were lowered to 50 lx and 65 lx by using neutral density filters placed on the screens. Before any testing took place, the mice were dark-adapted in their home cages with free access to food and water.

**Visual acuity.** A mouse was placed on top of the platform and allowed to acclimatize for a short period of time. Testing began when the mouse was no longer actively moving around. The visual acuity threshold was determined with contrast set at 100% and a grating of low spatial frequency (0.042 cycles per degree) as previously described in a recent report. When tracking behaviour was observed, the same stimuli were rotated in the anticlockwise direction (thus effectively testing the right eye). A staircase method of determining acuity threshold was implemented. A series of gratings of increasingly higher spatial frequencies were presented (rotating in one, then in the alternative direction) as long as the mouse indicated that it could detect the grating movements. When the mouse ceased to respond to a particular spatial frequency, a lower frequency grating was presented; when the mouse responded to a frequency, the frequency was increased. The acuity threshold was set as the highest spatial frequency to which the animal responded.

**Contrast-sensitivity function.** A contrast threshold was measured for six spatial frequencies (0.031, 0.064, 0.092, 0.103, 0.192, 0.272 cycles per degree). The initial contrast was set at 100% for each of the above spatial frequencies. Contrast was lowered until the mouse ceased to respond to the particular grating. The lowest contrast that elicited a response was assessed for each of the six spatial frequencies, and a contrast-sensitivity function was calculated with the formula 100/C, where C is the lowest contrast that elicits a response at a particular frequency. This data transform means that when a mouse can see at a very low contrast setting, the sensitivity number will be large and will indicate better visual performance at a particular spatial frequency.

## Immunofluorescence and laser scanning confocal microscopy

For confocal microscopy, converted and sorted cells were seeded on a chambered coverglass coated with 0.2% gelatin (Nunc). For mROS detection, cells were stained with MitoTracker Red (500 nM; Molecular Probes, M7512) for 30 min, washed, and used for downstream applications. For antibody staining, cells were fixed with 4% paraformaldehyde for 15 min and permeabilized with 0.25% Triton. Cells were then stained with primary antibody overnight at 4° C. After incubation with the appropriate secondary antibody, images were captured using a Zeiss LSM 510 confocal microscope. Data analysis and 3D reconstruction were performed with the assistance of ZEN lite software. Alexa Fluor 633-, 549- and 488-tagged secondary antibodies were used. For counting *Nrl*-GFP<sup>+</sup> CiPCs, ten 20× visual fields were selected randomly. The list of primary antibodies is provided in Supplementary Table 4.

## Measurement of mitochondrial ROS and TNF, H<sub>2</sub>O<sub>2</sub> treatment during reprogramming

Mitochondrial ROS was detected and quantified using a published protocol<sup>47</sup>. In brief, CiPCs, *Axin2*-depleted MEFs or reprogramming intermediates were incubated with MitoSOX Red (500 nM; Molecular Probes, M36008) for 30 min. Cells were then washed twice with 1× PBS and fluorescence was monitored with a microplate reader set to an excitation wavelength of 510 nm (excitation bandwidth, 10 nm) and an emission wavelength of 595 nm (emission bandwidth, 35 nm). mROS generation was also imaged with a fluorescence microscope and quantified with Leica Application Suite X Software. The region of interest was outlined for each cell on the image after background subtraction. The average intensity within each region of interest was measured and exported to an Excel spreadsheet. The average change in fluorescence was calculated for each type of cell. There were at least three replicates for each condition.



For H<sub>2</sub>O<sub>2</sub> production during reprogramming, D-galactose (0.5 mM) and galactose oxidase (0.015, 0.05, 0.1, 1 U ml<sup>-1</sup>) were added on day 4 of reprogramming along with all the small molecules<sup>48</sup>. On day 10–11, the number of CiPCs was quantified under the microscope (20 fields were randomly selected for counting). For TNF $\alpha$  treatment, cells were treated with three different concentrations of TNF $\alpha$  (20 ng ml<sup>-1</sup>, 50 ng ml<sup>-1</sup>, 100 ng ml<sup>-1</sup>) on day 8 and the number of cells was counted on day 11.

#### Measurement of oxygen consumption rate and extracellular acidification rate by Seahorse assay

About 25,000 CiPCs in 100  $\mu$ l medium with all small molecules were seeded in XF24 cell culture plates. Blank medium was added to the appropriate wells for background correction. Plates were then placed in an incubator at 37 °C with 5% CO<sub>2</sub> for 6 h. This step enables the cells to adhere on the surface of the plate. Next, 150  $\mu$ l growth medium containing all small molecules was added to each well followed by incubation overnight at 37 °C and 5% CO<sub>2</sub>. The following morning, the experiment was performed in a Seahorse Extracellular Flux Analyzer according to the manufacturer's instruction at the Metabolic Phenotyping Core facility, University of Texas South Western Medical Center.

#### FACS

For FACS, CiPCs were trypsinized (0.25%) and passed through a 40- $\mu$ m Nylon cell strainer (Fisher Scientific, Cat# 08-777-1) and suspended in PBS containing 3% bovine serum. Initiating MEFs were used as a negative control. Cells were then sorted in a Beckton-Dickinson LSRII Flow cytometer at the Flow Cytometry core facility, University of Texas Southwestern Medical Center and a Sony SH800 cell sorter at Flow Cytometry core facility, University of North Texas Health Science Center. Sorted cells were collected in IMR90 medium, centrifuged, and processed for RNA extraction and other downstream applications.

#### RNAi and generation of shRNA transduced MEFs

Lentiviral doxycycline-inducible shRNA constructs for *Axin2* (GE Dharmacon, 12006), *Tfam* (GE Dharmacon, 21780) and *RelA* (Sigma, TRCN00023583) were purchased for lentivirus preparation. Lentiviral supernatants were collected for 4 days and concentrated using lenti-X concentrator (Clontech, Cat# 631231). An aliquot of concentrated lentivirus was then used to transduce PO *Nrl*-GFP MEFs. For *Ascl1* knockdown experiments, control (scramble) shRNA and lentiviral shRNA (a gift from J. Johnson, UT-Southwestern Medical Center; cloned into PLKO.1 (Addgene)) constructs were transduced in MEFs. All lentiviral-transduced cells were selected for 3d in the presence of puromycin (1  $\mu$ g ml<sup>-1</sup>). Drug-selected cells were then used for chemical conversion and shRNA induction (*Tfam* and *Axin2*) was performed between day 4 and day 8 with doxycycline. Finally, CiPCs were quantified on day 11.

#### Western blotting

Whole-cell lysates were prepared as follows: cells were washed twice with ice-cold PBS, then lysed in RIPA buffer (Thermo Fisher Scientific) with 1X protease and phosphatase inhibitor cocktail (Thermo Fisher Scientific) on ice for 15–20 min. The mitochondrial and cytosolic fractions were prepared using a commercial kit (Thermo Fisher Scientific, Cat# 89874). Protein concentration was measured using a Pierce BCA protein assay kit (Thermo Fisher Scientific), and optical density at 562 nm was measured using a Biotek Synergy 2 Microplate Reader (BioTek Instruments). The NuPAGE protein gel system was used to separate cell lysates by electrophoresis. Blots were incubated in appropriate primary antibodies overnight and incubated with western HRP substrate (Millipore) or SuperSignal West Femto Maximum Sensitivity Substrate (34094). Images were acquired using the Innotech FluorChem HD2 imaging system (Alpha Innotech).

#### ChIP and amplification

Real-time PCR-based quantitative ChIP analysis was performed using a ChIP Assay Kit according to the manufacturer's instructions (EMD Millipore, Cat# 17-295). In brief, 20,000 cells were cross-linked with formaldehyde (to a final concentration of 1%) for 10 min at room temperature with gentle agitation. Cell sonication was performed such that the size of the chromatin would be between 300–500 bp. After pre-clearance with protein A agarose, chromatin was used for immunoprecipitation with specific antibodies targeted to NF- $\kappa$ B-p65. Immunoprecipitated chromatin was amplified with a Whole-Genome Amplification Kit (Sigma, WGA2). Amplified products were identified in agarose gel. Primers were designed to amplify 60–100 bp amplicons and were based on sequences from the Ensembl Genome Browser for mouse. Products were amplified with Fast SYBR Green Master Mix in a 20- $\mu$ l reaction. The amount of product was determined relative to a standard curve of input chromatin. Dissociation curves showed a single product for the amplicons. Primers for ChIP analysis are detailed in Supplementary Table 3.

#### BrdU labelling and staining

BrdU (1  $\mu$ M) was added on day 3 of reprogramming, and conversion was continued until day 11. CiPCs were washed twice with 1 $\times$  PBS and immunofluorescence was performed according to the published procedure (<https://www.abcam.com/protocols/brdu-staining-protocol>).

#### Preparation of *Nrl*-DsRed promoter reporter cells

To prepare the *Nrl*-DsRed promoter reporter, we digested out the promoter and reporter fragment from a commercially available vector, p*Nrl*-DsRed (Addgene, 13764), using restriction enzymes. These fragments were then cloned into a gateway entry vector pENTR2B (Thermo Fisher Scientific, A10463). Positive clones were then shuffled into a destination vector, pLentiX1 Zeo DEST (Addgene, 17299). The final product was then used for lentivirus preparation. Human fetal lung fibroblasts (HFL1; ATCC, CCL153) were transduced with the lentivirus and selected with Zeocin (200  $\mu$ g ml<sup>-1</sup>; InvivoGen, ant-zn-1) for 8 days.

#### Statistical analysis

All data are presented as mean  $\pm$  s.e.m. Statistical significance was determined using Student's *t*-test and one-way ANOVA using GraphPad Prism Software (GraphPad Software); *P* values are indicated in the figures.

#### Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

#### Data availability

Supporting RNA-seq and ATAC-seq data are deposited in the Gene Expression Omnibus under accession codes GSE138520 (RNA-seq) and GSE138521 (ATAC-seq) respectively. Source data for Figs. 1–4 and Extended Data Figs. 1–3, 5–8 are available within the manuscript files.

44. Jozefczuk, J., Drews, K. & Adjaye, J. Preparation of mouse embryonic fibroblast cells suitable for culturing human embryonic and induced pluripotent stem cells. *J. Vis. Exp.* **3854** (2012).
45. Chen, H. Y. et al. Three-dimensional retinal organoids from mouse pluripotent stem cells mimic in vivo development with enhanced stratification and rod photoreceptor differentiation. *Mol. Vis.* **22**, 1077–1094 (2016).
46. Kaya, K. D. et al. Transcriptome-based molecular staging of human stem cell-derived retinal organoids uncovers accelerated photoreceptor differentiation by 9-*cis* retinal. *Mol. Vis.* **25**, 663–678 (2019).
47. Wojtala, A. et al. Methods to monitor ROS production by fluorescence microscopy and fluorometry. *Methods Enzymol.* **542**, 243–262 (2014).
48. Tormos, K. V. et al. Mitochondrial complex III ROS regulate adipocyte differentiation. *Cell Metab.* **14**, 537–544 (2011).

# Article

**Acknowledgements** S.H.C. is supported by the Nancy Lee and Perry R. Bass Endowment, Foundation Fighting Blindness, and NEI awards EY021171 and EY025667. A.S. is supported by NEI Intramural Research Program (ZIAEY000450, ZIAEY000474 and ZIAEY000546). S.M. and S.B. are supported by NEI awards EY025905 and EY025717. T.M. is supported by 2T32AG020494-16A1. We thank A. Quiambao for performing the in vivo rodent injection studies, A. Ganguly for pupillometry assistance, the Histology Research Core at the University of North Carolina at Chapel Hill, and L. Gieser of NNRL, NEI for performing RNA-seq.

**Author contributions** B.M., S.H.C., S.M., R.F., A.N. and M.J.F. designed experiments; B.M. and S.B. performed experiments; A.S. designed, interpreted and supervised RNA-seq experiments and analysis; K.D.K. performed RNA-seq data analysis; W.Z.. assisted with rodent experiments; Y.F. performed HADF conversion; N.S., R.A.S., D.D. and T.M. performed in vivo experiments for vision restoration; Z.H. performed subretinal injection for some of the rodent experiments; and B.M., S.H.C. and A.S. wrote the manuscript.

**Competing interests** B.M. and S.H.C. are listed as inventors on a pending patent application (PCT/US2018/037784), filed by the University of North Texas, that covers the chemical reprogramming method reported in this paper. S.H.C. owns equity in and is a founder of CIRC Therapeutics, Inc., a start-up company that plans to commercialize novel treatments for ocular disease related to this technology. All other authors declare no competing interests.

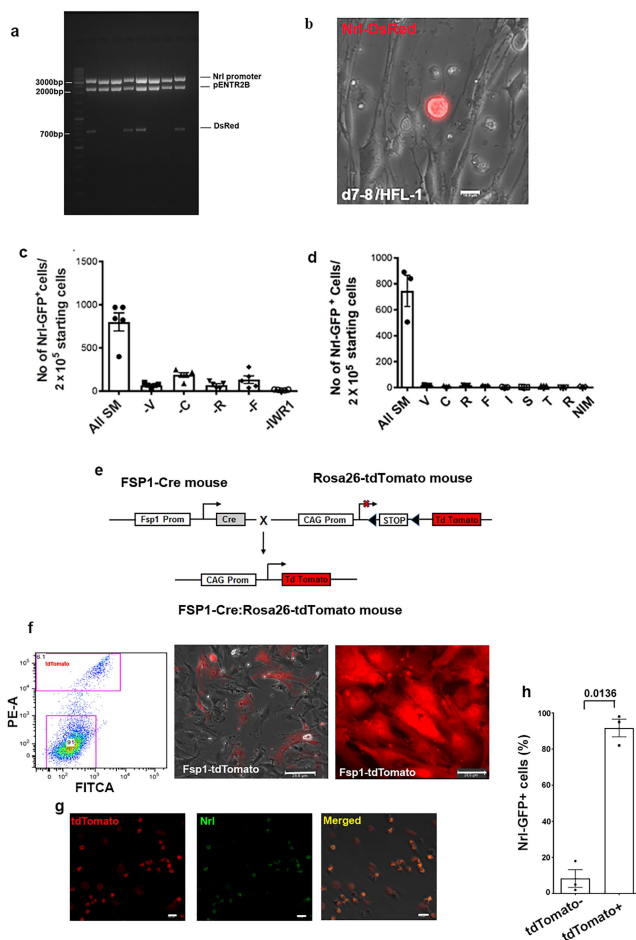
## Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41586-020-2201-4>.

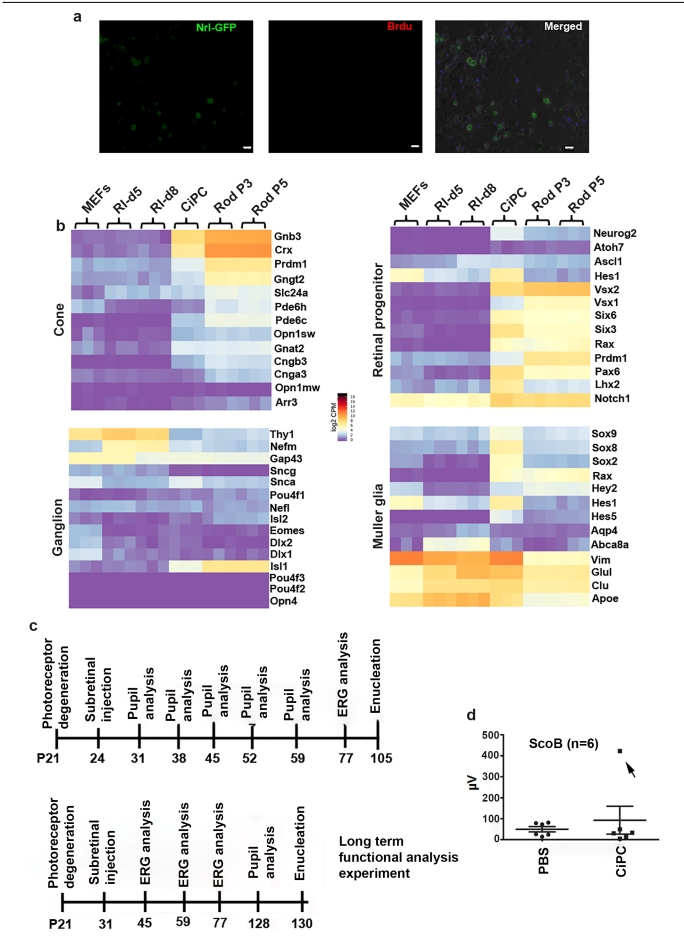
**Correspondence and requests for materials** should be addressed to S.H.C.

**Peer review information** *Nature* thanks Jayakrishna Ambati, Pete Coffey and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>.

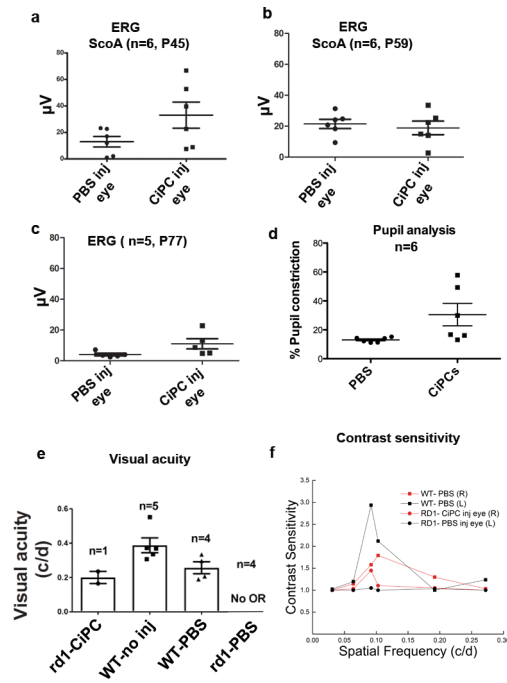


**Extended Data Fig. 1 | Preparation of *Nrl*-DsRed lentiviral reporter construct and lineage tracing.** **a**, DNA gel shows cloned *Nrl* and *DsRed* fragments digested out from the pENTR2B vector. This experiment was performed once. **b**, Expression of *Nrl*-DsRed in human CiPCs on day 8. **c**, The number of *Nrl*-GFP<sup>+</sup> cells upon subtraction of the indicated small molecules from the mixture. **d**, The number of *Nrl*-GFP<sup>+</sup> cells after conversion (on day 11) upon treatment with each small molecule alone. **e**, Scheme showing the generation of the mouse model for lineage tracing. **f**, Flow sorting of FSP1-Cre-tdTomato<sup>+</sup> MEFs. Left, scatter plot from FACS; middle, MEFs before cell sorting; right, MEFs after cell sorting. Scale bars, 25  $\mu$ m. This experiment was performed once. **g**, Cells expressing NRL are FSP1-tdTomato-positive on day 11, suggesting that they originate from fibroblasts. **h**, Percentage of cells that are positive for both NRL and tdTomato, compared with those that are just positive for NRL. In **c**, **d**, **h** data are presented as mean  $\pm$  s.e.m. of  $n = 3$  independently treated wells. Statistical significance was assessed using a two-tailed Student's *t*-test.

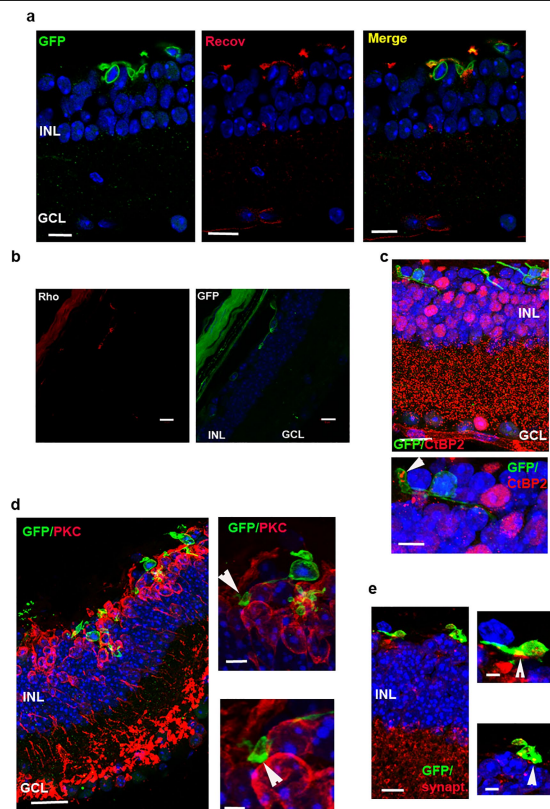


**Extended Data Fig. 2 | BrdU staining, transcriptome analysis and functional testing of CiPCs.** **a**, Images of BrdU-stained CiPCs on day 11. Scale bars, 10  $\mu$ m. **b**, Heat map analysis of RNA-seq data for the indicated cell-type-specific genes. **c**, Timeline for subretinal transplantation of CiPCs and functional analysis. **d**, Scotopic b-wave after transplantation of converted CiPCs. *n* denotes the number of biologically independent animals.



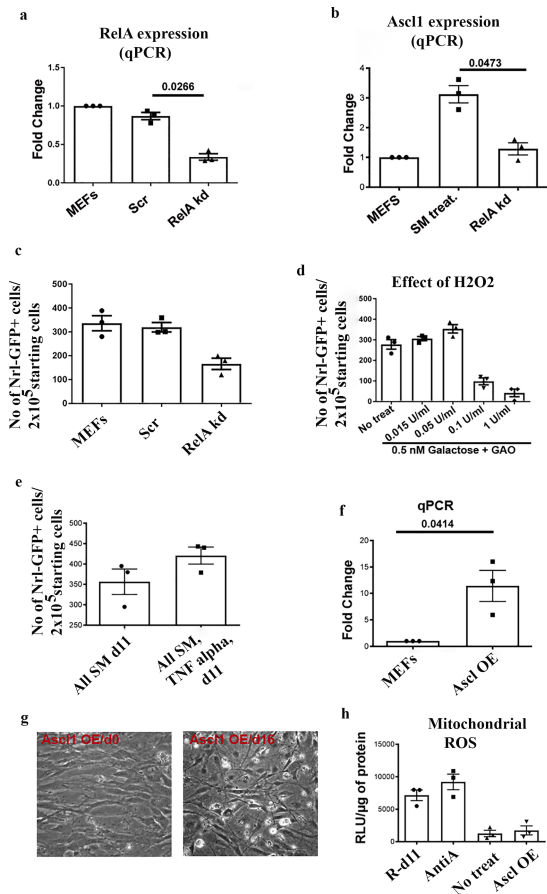


**Extended Data Fig. 3 | Functional analysis CiPCs in retinal degeneration mice (*rd1*).** **a–c**, Analysis for the scotopic a-wave of ERG after transplantation of CiPCs at day 45 (**a**), day 59 (**b**) and day 77 (**c**). **d**, Scatter plot of pupil constriction measurements for CiPC-injected *rd1* mice taken three months after transplantation. **e, f**, Measurement of visual acuity (**e**) and contrast sensitivity (**f**) for CiPC-injected mice that demonstrated improved pupil response after transplantation and control mice. OR, optokinetic response. For the CiPC-injected *rd1* experiment, repeated values are plotted (for  $n = 1$  mouse). In **a–e**,  $n$  denotes biologically independent animals and the experiment in **f** was performed twice.

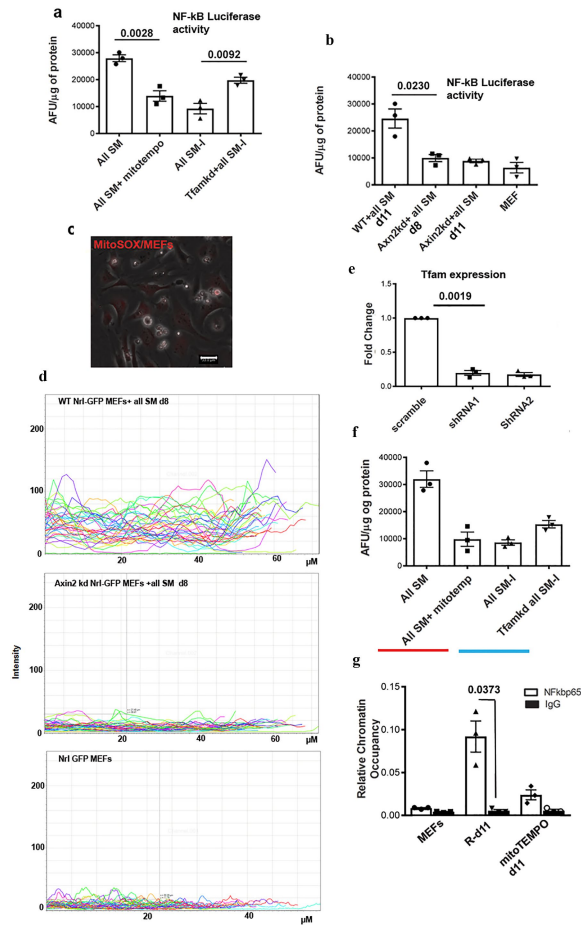


**Extended Data Fig. 4 | Transplanted CiPCs express photoreceptor-specific markers and connected to the inner retina three months after transplantation.**

**a**, Expression of recoverin (Recov, red) in CiPCs (green). **b**, Expression of rhodopsin (Rho, red) in GFP-expressing CiPCs. **c**, Expression of rod ribbon-synapse protein ribeye (red) in transplanted CiPCs (green). The bottom panel shows a magnified version of the image (arrowhead). **d**, Close apposition of transplanted CiPCs (green) with PKC-positive (red) rod bipolar cells (white arrowhead). **e**, Expression of synaptic marker protein synaptophysin (red, arrowhead) in transplanted CiPCs (green) (co-localization, yellow). **d, e**, Left, images at 20× magnification; right, images at higher digital magnification. In **a–e**, experiments were repeated twice with similar results. Scale bars, 10 μm.

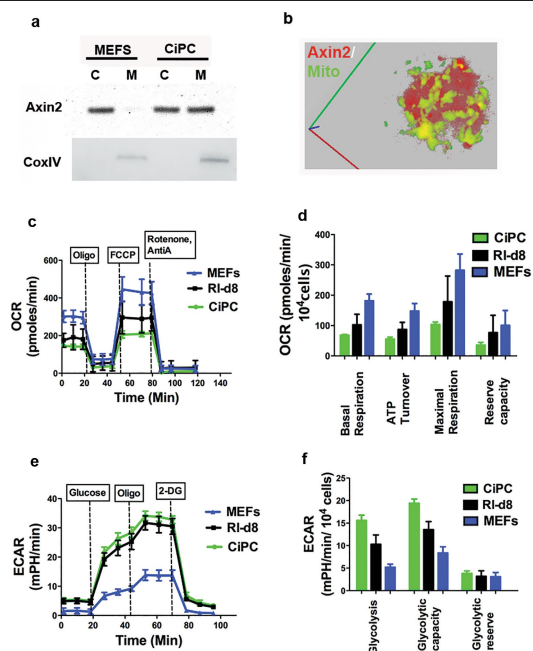


**Extended Data Fig. 5 | Effect of NF- $\kappa$ B knockdown, H<sub>2</sub>O<sub>2</sub>, TNF $\alpha$  and the overexpression of *Ascl1* on the reprogramming of fibroblasts to CiPCs.**  
**a**, qPCR analysis following shRNA-mediated knockdown (kd) of *RelA* (which encodes NF- $\kappa$ B) in MEFs. **b**, qPCR analysis of *Ascl1* expression after the reprogramming of *RelA*-depleted MEFs on day 11. **c**, Number of *Nrl*-GFP cells after the conversion of *RelA*-knockdown MEFs on day 11. **d**, Effect of different concentrations of H<sub>2</sub>O<sub>2</sub>, generated by galactose and galactose oxidase (GAO), on the reprogramming of fibroblasts to CiPCs. **e**, Effect of cytokine (TNF, 50 ng ml<sup>-1</sup>) on CiPC conversion. **f**, qPCR analysis showing the overexpression (OE) of *Ascl1* in MEFs. **g**, Phase contrast micrographs of MEFs overexpressing *Ascl1* before and after reprogramming in inducing medium. No *Nrl*-GFP<sup>+</sup> cells were evident up to day 15. **h**, Fluorometric measurement of mROS in *Ascl1*-overexpressing cells on day 8 of reprogramming. Data are presented as mean  $\pm$  s.e.m. of  $n = 3$  independently treated wells. Statistical significance was assessed using a two-tailed Student's *t*-test.

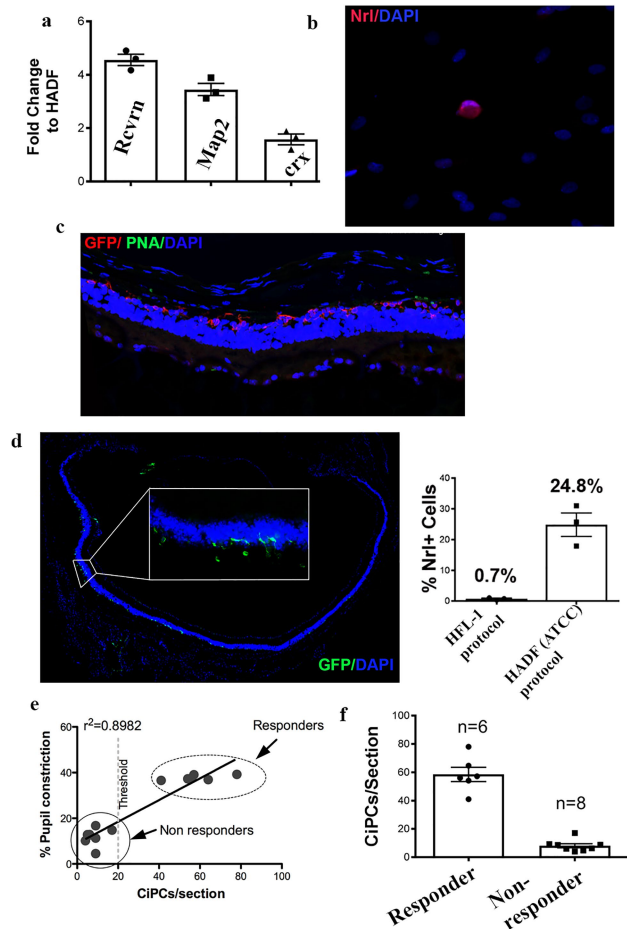


**Extended Data Fig. 6 | NF-κB-luciferase activity and generation of mROS during the reprogramming of fibroblasts to CiPCs.** **a**, A luciferase activity assay reveals decreased activation of NF-κB upon treatment with MitoTEMPO and increased NF-κB activity upon depletion of TFAM on day 11 of reprogramming to CiPCs. **b**, Luciferase activity shows decreased NF-κB activation in *Axin2*-knockdown MEFs on day 8 and day 11 of reprogramming. SM, small molecules. **c**, MEFs stained with MitoSOX showed low levels of mROS production (scale bar, 33 μm). **d**, Plot of mROS accumulation on day 8, as visualized using fluorescence microscopy after MitoSOX staining, in wild-type (WT) *Nrl*-GFP MEFs treated with all compounds (top), *Axin2*-knockdown *Nrl*-GFP MEFs treated with all compounds (middle) and control *Nrl*-GFP MEFs (bottom). **e**, qPCR analysis following *Tfam* depletion in MEFs using shRNA. **f**, Fluorometric measurement of mROS (after MitoSOX staining) in MitoTEMPO-treated and TFAM-knockdown cells on day 11. **g**, ChIP assay shows reduced binding of activated NF-κB near the *Ascl1* locus upon treatment with MitoSOX. In **a**, **b**, **e**–**g**, data are presented as mean ± s.e.m. of  $n = 3$  independently treated wells. Statistical significance was assessed using a two-tailed Student's *t*-test.

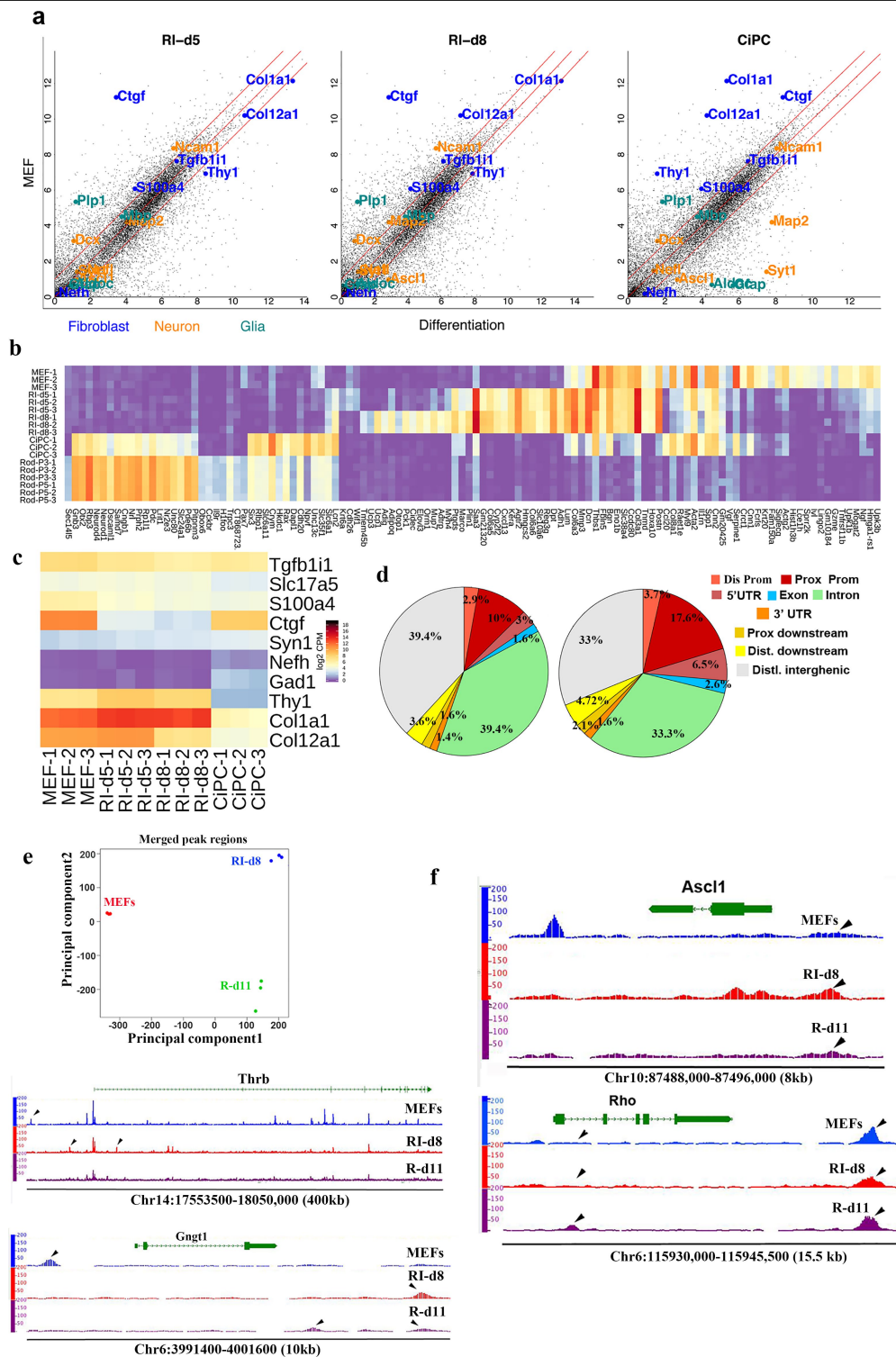




**Extended Data Fig. 7 | Mitochondrial localization of AXIN2 in converted cells and a mito-stress test for CiPCs.** **a**, Western blot after subcellular fractionation of CiPCs showed that AXIN2 was localized in the mitochondria. C, cytoplasm; M, mitochondria. For gel source data, see Supplementary Fig. 1e. **b**, Micrograph of AXIN2 (red)- and GFP (green)-stained CiPCs after z-stack and 3D reconstruction, also showing that AXIN2 localizes in the mitochondria. In **a**, **b** experiments were performed once. **c**, CiPCs were found to have a reduced oxygen consumption rate (OCR) compared with MEFS and day-8 reprogramming intermediates, as measured using a Seahorse assay. Oligo, oligomycin; FCCP, carbonyl cyanide-4-phenylhydrazone. **d**, The indicated mitochondrial parameters of CiPCs, quantified in terms of OCR using a Seahorse assay, compared with those of MEFS and day-8 reprogramming intermediates. **e**, The extracellular acidification rate (ECAR) of CiPCs, MEFS and day-8 reprogramming intermediates was measured using a Seahorse assay. **f**, The indicated mitochondrial parameters of CiPCs, quantified in terms of ECAR using a Seahorse assay, compared with those of MEFS and day-8 reprogramming intermediates. In **c**–**f**, data are presented as mean  $\pm$  s.e.m. of  $n = 3$  independently treated wells.



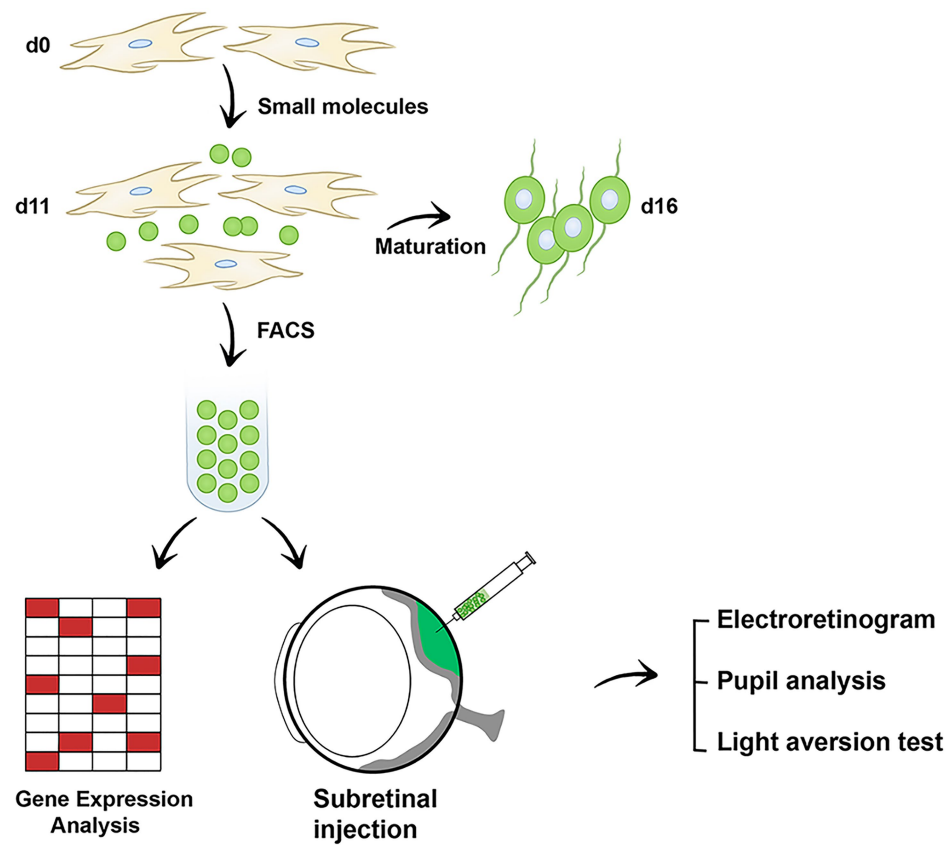
**Extended Data Fig. 8 | Conversion of human fibroblasts into CiPCs and in vivo functional testing of mouse CiPCs. a**, qPCR analysis shows expression of the indicated photoreceptor-specific genes in CiPCs. **b**, Fluorescence microscopy image demonstrates the expression of NRL in CiPCs. **a, b**, HADF from patients with retinitis pigmentosa. This experiment was performed three times with similar results. **c**, Confocal microscopy image of the retina of an *rd1* mouse, immunostained with GFP (red) and PNA (peanutagglutinin, green), three months after transplantation of CiPCs. **d**, Left, confocal microscopy image of a retinal section showing GFP-positive CiPCs three months after transplantation. This experiment was repeated twice with similar results. Right, conversion of human HFL1s (original protocol, see Methods) and a modified protocol using HADFs (Fig. 1f) to CiPCs. **e**, The correlation between the number of transplanted CiPCs and the extent of pupil constriction.  $n = 6$  eyes classified as pupil responders and  $n = 8$  eyes classified as pupil non-responders. **f**, CiPC survival in the eyes of *rd1* mice classified as pupil responders and pupil non-responders. The  $n$  values denote the number of eyes. In **a, d, e**, data are presented as mean  $\pm$  s.e.m. of  $n = 3$  independently treated wells.



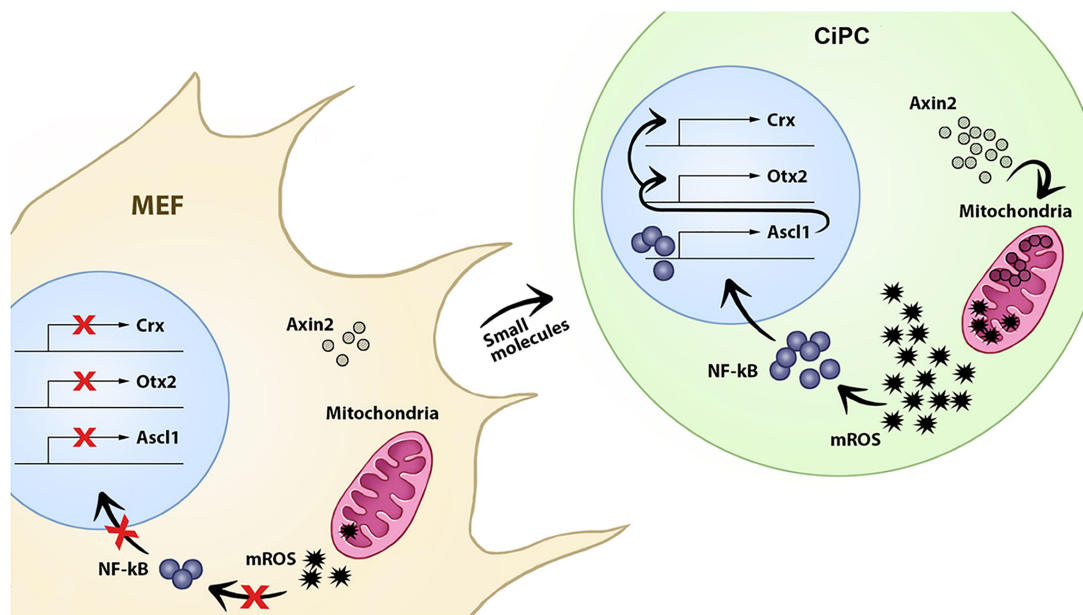
**Extended Data Fig. 9 | Gene expression and chromatin analysis of CiPCs.**  
**a**, Pairwise gene expression comparison of RNA-seq data demonstrates that neuronal markers are increased, glial markers are not activated and fibroblast genes are silenced. Numbers on both axes represent logarithmic base 2-transformed counts per million values for each gene. **b**, Heat map analysis (from RNA-seq data) of photoreceptor genes along with other neuronal and fibroblast genes. **c**, Fibroblast-specific gene expression as determined by RNA-seq during reprogramming. The expression of *Col1a1*, *Col2a1*, *Thy1*, *Ctgf*,

*S100a4* (also known as *Fsp1*) and *Slc17a5* is reduced in reprogramming intermediates as well as in CiPCs. **d**, Pie charts of results of ATAC-seq analysis show increased open chromatin corresponding to the proximal promoter region in reprogramming intermediates. **e**, Principal component analysis of ATAC-seq data shows open chromatin regions for the indicated samples. **f**, ATAC-seq analysis shows opening of specific chromatin regions for the indicated photoreceptor-specific loci (*Ascl1*, *Rho*, *Gngt1* and *Thrb*). All data are representative of three independently treated samples.

a



b



**Extended Data Fig. 10 | Schematic of the transplantation study and the mechanism of reprogramming fibroblasts to CiPCs. a,** Schematic of reprogramming to CiPCs and functional analysis. **b,** Schematic of the mechanism of reprogramming fibroblasts to CiPCs.



## Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

### Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- ☐ ☒ The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement
- ☐ ☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- ☐ ☒ The statistical test(s) used AND whether they are one- or two-sided  
*Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
- ☒ ☐ A description of all covariates tested
- ☒ ☐ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- ☐ ☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- ☐ ☒ For null hypothesis testing, the test statistic (e.g.  $F$ ,  $t$ ,  $r$ ) with confidence intervals, effect sizes, degrees of freedom and  $P$  value noted  
*Give  $P$  values as exact values whenever suitable.*
- ☒ ☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- ☒ ☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- ☒ ☐ Estimates of effect sizes (e.g. Cohen's  $d$ , Pearson's  $r$ ), indicating how they were calculated

*Our web collection on [statistics for biologists](#) contains articles on many of the points above.*

### Software and code

Policy information about [availability of computer code](#)

Data collection Stepone V2.3, Applied Biosystems; LASX, Leica; ZEN2.3SP1, Zeiss; ZEN2.3, Zeiss; Enduro GDS, Labnet International.

Data analysis ImageJ (NIH), ZEN lite, Leica Application, Suite X, GraphPad Prism v5.0, Origin2019

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

### Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Supporting RNASeq and ATACSeq data are deposited in the in the Gene Expression Omnibus under accession code GSE138520 (RNASeq) and GSE138521 (ATACSeq) respectively. Source data for Fig. 1-4 and Extended Data Fig. 1-3,5-8 are available within the manuscript files.

## Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

- ☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

## Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Sample size were chosen based upon previously described studies. No statistical methods were used to predetermine sample sizes.
Data exclusions	No data were excluded.
Replication	Each experiment had at least three replicates, and was performed two or more times.
Randomization	Animals were randomly assigned to groups.
Blinding	Investigators were masked to group identity for light aversion experiments.

## Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems		Methods	
n/a	Involved in the study	n/a	Involved in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies	<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines	<input type="checkbox"/>	<input checked="" type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology	<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging
<input type="checkbox"/>	<input checked="" type="checkbox"/> Animals and other organisms		
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants		
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data		

### Antibodies

Antibodies used	Axin2/conductin, Goat, Santa Cruz, 1:100 (IF), 1:200 (WB) SC-8570; Nrl, Mouse monoclonal, Santa Cruz, 1:100(IF), SC-398046; Crx, Mouse monoclonal, Santa Cruz, 1:200 (IF), SC-377138; Beta actin, Mouse, Thermo Scientific, 1:2000 (WB), MA5-15739; GFP, Chicken, Abcam, 1:400 (IF), Ab13970; Axin1, Goat, RD systems, 1:400 (WB), AF3287; Rhodopsin, Mouse, Millipore, 1:100 (IF), MAB5316; Recoverin, Rabbit, Abcam, 1:100 (IF), Ab64945; ChX10, Sheep, Millipore, 1:100 (IF), AB9016; BrdU, Mouse, Thermo scientific, 1:500 (IF), MA3071; NF-kB Mouse Cell Signaling 6µg (ChiP) 6956P; Pkc, Mouse, Santa Cruz, 1:100(IF), SC-8393; Synaptophysin, Rabbit, Abcam, 1:500 (IF), Ab14692; CtBP2 (ribeye), Rabbit, BD, 1:500(IF), 612044.
Validation	All of the antibodies used in this study were used according to supplied instructions by the manufacturer or referenced publication.

### Animals and other organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research

Laboratory animals	For pupil reflex and light aversion experiments, CiPCs were injected in to P31 or P24 rd1 mice (Jackson Laboratory, 000656). Males and females were used approximately in an equal ratio. FSP1-Cre (Jackson Laboratory, 012641) and R26-LSL-tdTomato (Jackson Laboratory, Ai9) crossed embryos were used for a standard procedure for MEF preparation (Li et al., 2015). All studies and animal care were performed in accordance with relevant guidelines and regulations as approved by the animal care and use committee at University of North Texas Health Science Center and EyeCRO.
Wild animals	This study did not involve wild animals.
Field-collected samples	This study did not involve samples collected from the field.
Ethics oversight	All animal studies and animal care were performed in accordance with relevant guidelines and regulations and approved by the animal care and use committee at University of North Texas Health Science Center.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

## Flow Cytometry

### Plots

Confirm that:

- ☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
- ☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).
- ☒ All plots are contour plots with outliers or pseudocolor plots.
- ☒ A numerical value for number of cells or percentage (with statistics) is provided.

### Methodology

Sample preparation	Chemically reprogrammed cell population was PBS-washed, trypsinized (0.25%) and passed through a 40µM nylon cell strainer (Fisher Scientific, 08-777-1) and suspended in PBS containing 3% bovine serum.
Instrument	Beckton-Dickinson LSRII Flow cytometer, Sony SH800 cell sorter
Software	BD FACSDIVA, Sony SH800, Flowjo
Cell population abundance	Approx 0.2% CiPCs were obtained from mouse embryonic fibroblasts and 0.9% were obtained from human fetal lung fibroblasts.
Gating strategy	Cells were first identified by FSC/SSC so that all cells are visible on the plot. FSC-H and FSC-A were used to identify single cells. Positive cell population was sorted out based upon green/red fluorescence expression.

- ☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.

# Plant 22-nt siRNAs mediate translational repression and stress adaptation

<https://doi.org/10.1038/s41586-020-2231-y>

Received: 2 August 2019

Accepted: 25 February 2020

Published online: 29 April 2020

 Check for updates

Huihui Wu<sup>1,2,8</sup>, Bosheng Li<sup>1,8</sup>, Hiro-oki Iwakawa<sup>3,4</sup>, Yajie Pan<sup>1,5</sup>, Xianli Tang<sup>1,6</sup>, Qianyan Ling-hu<sup>1</sup>, Yuelin Liu<sup>1</sup>, Shixin Sheng<sup>1</sup>, Li Feng<sup>1</sup>, Hong Zhang<sup>1</sup>, Xinyan Zhang<sup>2,7</sup>, Zhonghua Tang<sup>5</sup>, Xinli Xia<sup>6</sup>, Jixian Zhai<sup>1</sup> & Hongwei Guo<sup>1✉</sup>

Small interfering RNAs (siRNAs) are essential for proper development and immunity in eukaryotes<sup>1</sup>. Plants produce siRNAs with lengths of 21, 22 or 24 nucleotides. The 21- and 24-nucleotide species mediate cleavage of messenger RNAs and DNA methylation<sup>2,3</sup>, respectively, but the biological functions of the 22-nucleotide siRNAs remain unknown. Here we report the identification and characterization of a group of endogenous 22-nucleotide siRNAs that are generated by the DICER-LIKE 2 (DCL2) protein in plants. When cytoplasmic RNA decay and DCL4 are deficient, the resulting massive accumulation of 22-nucleotide siRNAs causes pleiotropic growth disorders, including severe dwarfism, meristem defects and pigmentation. Notably, two genes that encode nitrate reductases—*NIA1* and *NIA2*—produce nearly half of the 22-nucleotide siRNAs. Production of 22-nucleotide siRNAs triggers the amplification of gene silencing and induces translational repression both gene specifically and globally. Moreover, these 22-nucleotide siRNAs preferentially accumulate upon environmental stress, especially those siRNAs derived from *NIA1/2*, which act to restrain translation, inhibit plant growth and enhance stress responses. Thus, our research uncovers the unique properties of 22-nucleotide siRNAs, and reveals their importance in plant adaptation to environmental stresses.

Small RNAs (sRNAs) are produced by different proteins, have different lengths and exert their effects on the cell through different mechanisms. In general, cellular double-stranded RNA (dsRNA) precursors that are produced by RNA-dependent RNA polymerases and hairpin microRNA (miRNA) precursors are cleaved into 20–24-nucleotide (nt) sRNA duplexes by proteins of the Dicer family<sup>1,3</sup>. One strand of the sRNA duplex is loaded into an Argonaute (AGO) protein to form an RNA-induced silencing complex (RISC)<sup>3</sup>. The *Arabidopsis* genome contains four Dicer-like (DCL) proteins, DCL1 to DCL4, with a variety of roles. DCL1 is the major DCL protein required for the biogenesis of miRNAs that target homologous RNAs for cleavage and/or translational repression<sup>4</sup>; DCL3 cleaves dsRNA precursors into 24-nt siRNAs, which are engaged in RNA-directed DNA methylation<sup>5</sup>; and DCL4 is required for the production of 21-nt siRNAs, directing mRNA cleavage<sup>6</sup>. DCL2 is thought to regulate the biogenesis of viral 22-nt siRNAs<sup>7</sup>. Although the roles of 21- and 24-nt siRNAs are well understood, the lack of endogenous 22-nt siRNAs has greatly restricted their study. Here, we explored the biological roles and regulatory mechanism of 22-nt siRNAs by using genetic mutants that produce abundant quantities of these molecules.

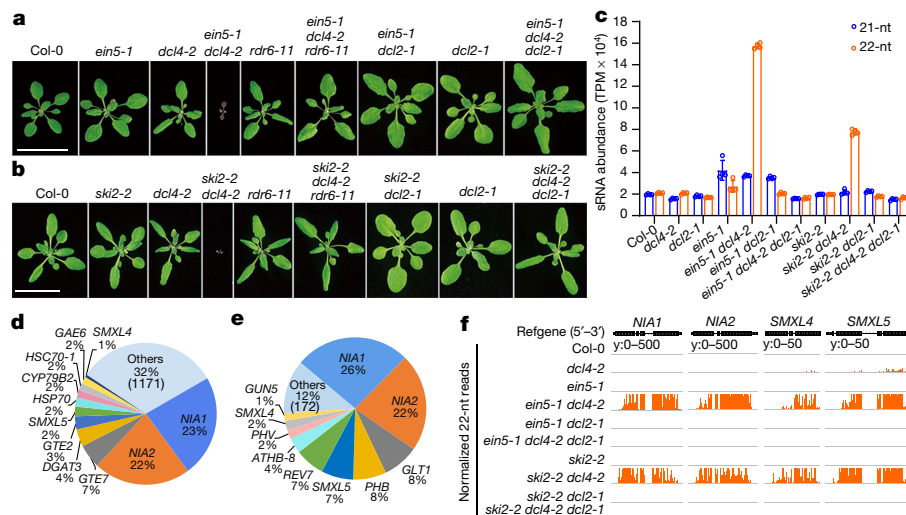
## Massive production of 22-nt siRNAs

In *Arabidopsis*, bidirectional (5′-to-3′ and 3′-to-5′) decay of cytoplasmic RNA, mediated by the ETHYLENE INSENSITIVE5/EXORIBONUCLEASE4 (EIN5/XRN4) and SUPER KILLER2 (SKI2) proteins, elaborately averts endogenous gene silencing by suppressing the accumulation of coding-transcript-derived siRNAs<sup>8</sup>. We began by investigating plants with mutations in these RNA-degrading enzymes and in DICER proteins. We found that simultaneous loss of cytoplasmic RNA decay and DCL4 activity (*ein5 dcl4* and *ski2 dcl4* double mutants) led to pleiotropic growth disorders, including severe defects in meristem activity, leaf expansion and pigmentation (Fig. 1a, b). By contrast, *ein5 dcl2* and *ski2 dcl2* mutants were normal. Interestingly, *dcl2* and a post-transcriptional gene silencing mutant, *rdr6*, rescued the defects in the *ein5 dcl4* and *ski2 dcl4* mutants (Fig. 1a, b), suggesting that these disorders result from the production of DCL2- and RDR6-dependent 22-nt siRNAs.

By carrying out sRNA sequencing, we identified abundant 22-nt but not 21-nt siRNAs derived from both *TRANS-ACTING siRNA* (*TAS*) and non-*TAS* loci in *ein5 dcl4* and *ski2 dcl4* mutants (Fig. 1c and Extended Data Fig. 1a, b), consistent with the competing relationship between DCL4

<sup>1</sup>Department of Biology, Institute of Plant and Food Science, Southern University of Science and Technology, Shenzhen, China. <sup>2</sup>State Key Laboratory of Protein and Plant Gene Research, School of Life Sciences, Peking University, Beijing, China. <sup>3</sup>Institute for Quantitative Biosciences, The University of Tokyo, Tokyo, Japan. <sup>4</sup>Japan Science and Technology Agency (JST), Precursory Research for Embryonic Science and Technology (PRESTO), Saitama, Japan. <sup>5</sup>Key Laboratory of Plant Ecology, Northeast Forestry University, Harbin, China. <sup>6</sup>Beijing Advanced Innovation Center for Tree Breeding by Molecular Design, National Engineering Laboratory for Tree Breeding, College of Biological Sciences and Technology, Beijing Forestry University, Beijing, China. <sup>7</sup>Present address: Lingnan Guangdong Laboratory of Modern Agriculture, Genome Analysis Laboratory of the Ministry of Agriculture, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen, China. <sup>8</sup>These authors contributed equally: Huihui Wu, Bosheng Li. ✉e-mail: guohw@sustech.edu.cn





**Fig. 1 | Disruption of cytoplasmic RNA decay and DCL4 triggers massive production of 22-nt siRNAs and causes growth disorders.** **a, b,** Images of 20-day-old plants of the indicated wild-type (Col-0) and mutant genotypes. Scale bars, 2 cm. Three biological repeats were performed with similar results. **c,** Abundance of 21- and 22-nt siRNAs (TPM, tags per million) from all genes, excluding miRNAs and tasiRNAs. Data are presented as means  $\pm$  s.d.;  $n = 4$

and DCL2 in substrate processing<sup>6,9</sup>. In contrast with 21-nt trans-acting siRNAs (tasiRNAs)<sup>6</sup>, no phasing was detected in the TAS-derived 22-nt siRNAs (Extended Data Fig. 1c; ‘phasing’ means that the sRNAs were generated precisely in a head-to-tail arrangement, starting from a specific nucleotide). In *ein5 dcl4* and *ski2 dcl4* plants, we identified, respectively 1,182 and 182 protein-coding genes that produce 22-nt siRNAs, with 111 genes in common between the two conditions (Supplementary Tables 1, 2 and Extended Data Fig. 1d, e). Two genes encoding nitrate reductases, *NIA1* and *NIA2* (ref.<sup>10</sup>), were particularly important, contributing 45% and 48% to the total amount of 22-nt siRNAs in *ein5 dcl4* and *ski2 dcl4* plants, respectively (Fig. 1d–f). Among the top 20 loci, *SMAXI-LIKE 4* (*SMXL4*) and *SMXL5*, which are required for phloem development<sup>11</sup>, also showed abundant generation of 22-nt siRNAs in both mutants (3% in *ein5 dcl4* and 9% in *ski2 dcl4*) (Fig. 1d–f). From a set of *HD-ZIP/III* transcription-factor genes (*PHB*, *PHV*, *REV* and *ATHB-8*), we detected abundant quantities of 22-nt siRNAs in *ski2 dcl4* mutants (21% of the total amount of 22-nt siRNAs), but not in *ein5 dcl4* mutants (Fig. 1d, e and Extended Data Fig. 1e).

## 22-nt siRNAs repress mRNA translation

Next, we carried out transcriptome profiling to assess the effect of 22-nt siRNAs on gene expression. We found that only a small subset of the genes that generate 22-nt siRNAs showed decreased expression levels in *ein5 dcl4* and *ski2 dcl4* mutants (Fig. 2a, b), including *SMXL4* and *SMXL5* (Fig. 2c and Extended Data Fig. 2a). The vast majority, including *NIA1* and *NIA2*, showed unchanged or higher expression levels in *ein5 dcl4* and *ski2 dcl4* plants in both the total and the cytoplasmic mRNA fractions (Fig. 2c and Extended Data Fig. 2a–d). Nevertheless, *NIA1* and *NIA2* proteins were hardly detectable, even in the presence of MG132, an inhibitor of the 26S proteasome, arguing against the proteasomal degradation of the two proteins (Fig. 2d and Extended Data Fig. 2e, f). Meanwhile, mRNAs from polysome-associated *NIA1/2* (Fig. 2e, f, fractions 8–11) and another two genes that produce 22-nt siRNAs, *GLOBAL TRANSCRIPTION FACTOR GTE2/7* (*GTE2/7*) (Extended Data Fig. 2g), in *ein5 dcl4* plants were lower than in wild-type or in *ein5 dcl4 dcl2* plants, suggesting the translational repression of *NIA1/2* and *GTE2/7* mRNAs by the 22-nt siRNAs. Notably, the total amount of

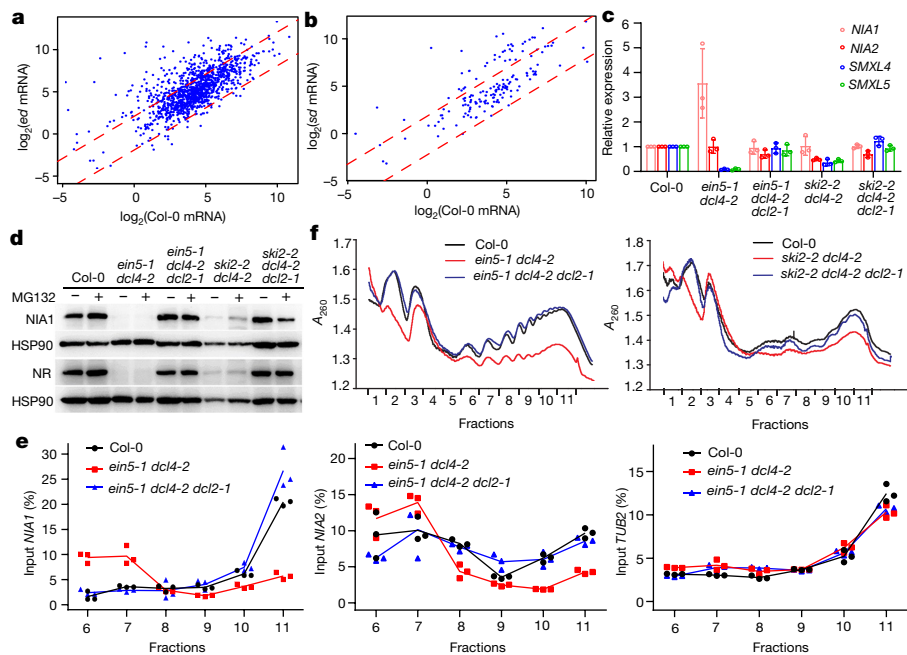
biologically independent samples. **d, e,** Percentages of top-scoring 22-nt siRNA-generating loci, ranked by 22-nt siRNA abundance in *ein5-1 dcl4-2* (**d**) and *ski2-2 dcl4-2* (**e**) mutants. **f,** Integrated Genome View (IGV) of 22-nt siRNAs from *NIA1/2* and *SMXL4/5* genes. y: 0–500 and y: 0–50 indicate the ranges of data values on the y-axis.

polysome-associated mRNAs (fractions 6–11) also decreased in *ein5 dcl4* and *ski2 dcl4* mutants; this was restored by the *dcl2* mutation, implying global translational repression (Fig. 2f). By comparison, despite the similar growth defects and abundant 21-nt siRNAs<sup>8</sup>, the total amount of polysome-associated mRNAs (fractions 6–11) in *ein5 ski2* mutants was decreased to a much lesser extent than in *ein5 dcl4* plants (Extended Data Fig. 3a–d). Furthermore, polysome-associated *NIA1* and *NIA2* mRNAs (fractions 6–11) in *ein5 ski2* mutants were comparable to those in the wild type (Extended Data Fig. 3e–g). Therefore, the production of 22-nt, but not 21-nt, siRNAs represses translation of their cognate mRNAs, as well as global translation.

## AGO1 is required for 22-nt siRNA activity

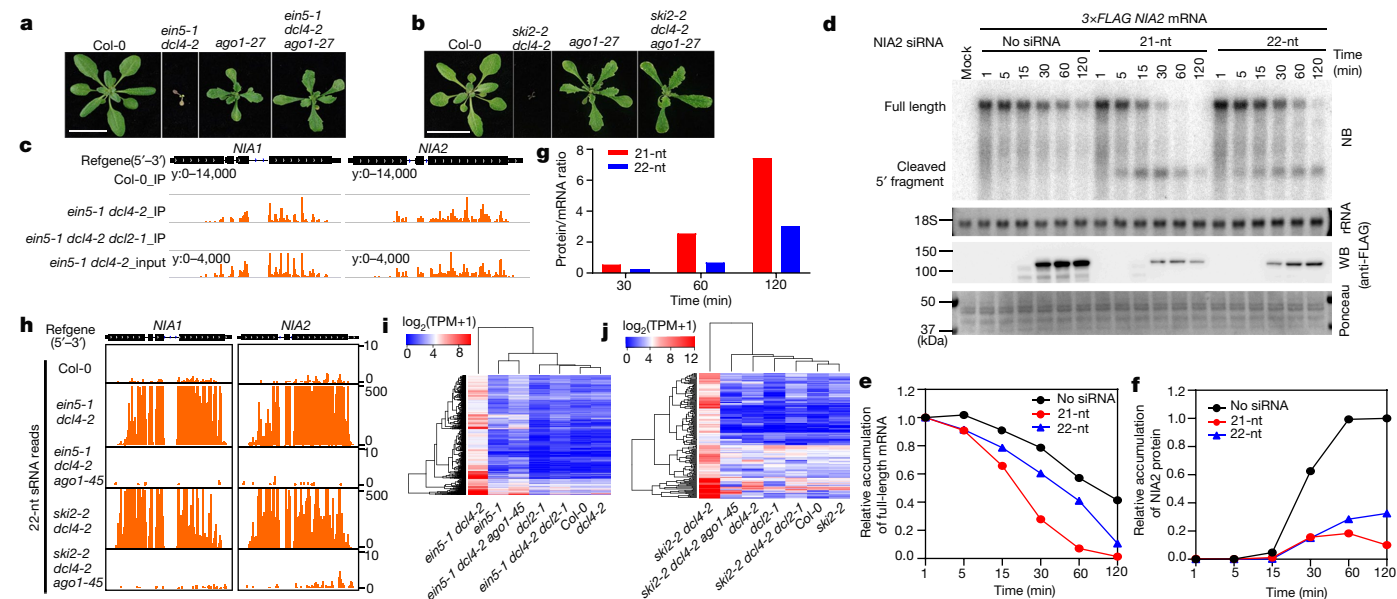
Further evidence supporting the ability of 22-nt siRNAs to repress translation came from the *ago1-27* mutant plant. This is a hypomorphic allele that has normal AGO1 ‘slicer’ activity but is defective in translational repression<sup>12,13</sup>. When *ago1-27* was crossed into *ein5 dcl4* and *ski2 dcl4* plants, the growth phenotype was rescued and protein levels of *NIA1/2* were restored (Fig. 3a, b and Extended Data Fig. 4a). This effect was specific to *ago1*: the *ago2* and *ago4* mutants could not rescue the growth phenotype (Extended Data Fig. 4b, c). *HUA ENHANCER 1* (*HEN1*) is required for sRNA methylation<sup>14</sup>; the *hen1* mutant partly rescued the defects of both *ein5 dcl4* and *ski2 dcl4* mutants (Extended Data Fig. 4d, e), suggesting that 22-nt siRNAs are methylated. However, *ALTERED MERISTEM PROGRAM 1* (*AMPI*), which is involved in miRNA-mediated translational repression<sup>15</sup>, was not required for 22-nt siRNA activity: the *ampi-30* mutation did not rescue *ein5 dcl4* and *ski2 dcl4* mutants (Extended Data Fig. 4f–h), suggesting that the translational repression mechanisms of miRNAs and 22-nt siRNAs are different.

Given the genetic evidence that AGO1 is required for 22-nt siRNA activity, we analysed AGO1-associated sRNAs by AGO1 immunoprecipitation followed by sRNA sequencing. AGO1 predominantly associated with 21-nt miRNAs in wild-type and *ein5 dcl4 dcl2* plants, but switched to binding 21-nt miRNAs and 22-nt siRNAs in *ein5 dcl4* mutants (Extended Data Fig. 5a, b). The AGO1-associated 21- and 22-nt sRNAs showed a 5′-uridine (U) preference (Extended Data Fig. 5c, d), consistent with previous findings<sup>16</sup>. Of 642 identified genes that produced



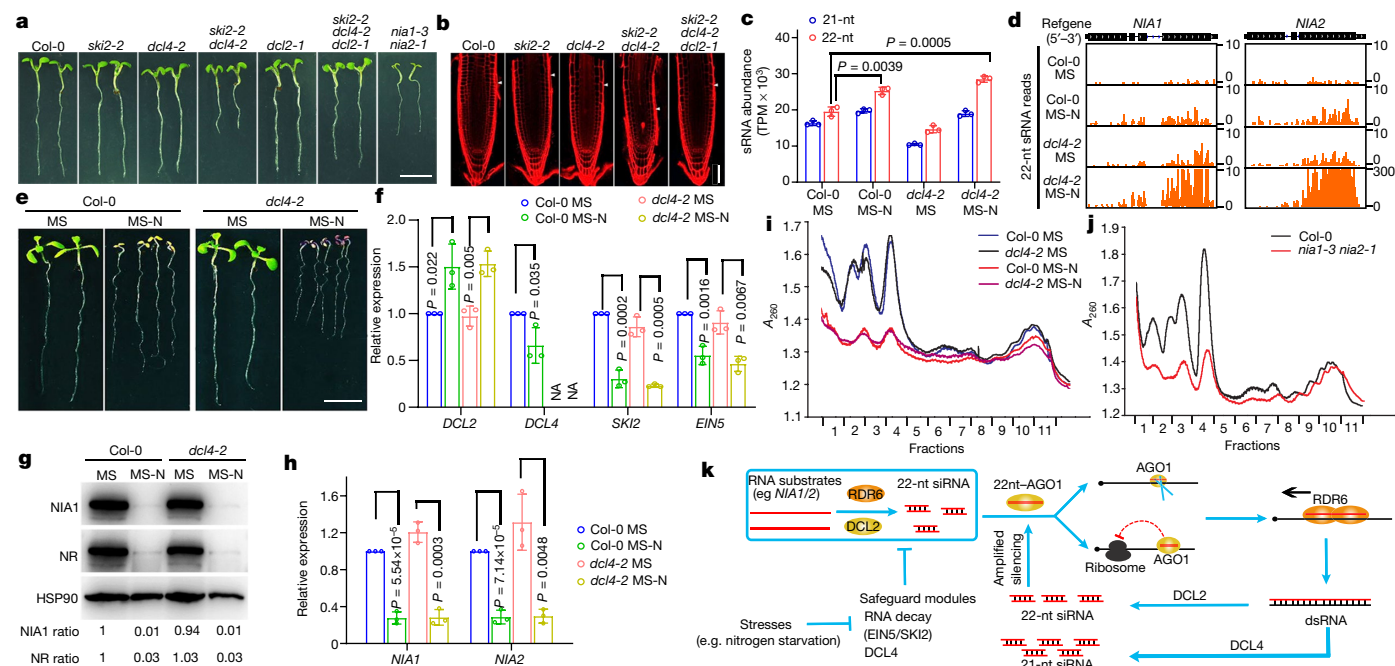
AGO1-associated 22-nt siRNAs, 529 (82%) overlapped with genes that generate 22-nt siRNAs in *ein5 dcl4* mutants (Extended Data Fig. 5e). The abundance of AGO1-associated 22-nt siRNAs positively correlated with that of total 22-nt siRNAs in *ein5 dcl4* plants (Extended Data Fig. 5f). As expected, large amounts of 22-nt siRNAs from *NIA1/2* were associated with AGO1 (Fig. 3c). Therefore, 22-nt siRNAs with a 5'-terminal U can be selectively loaded into AGO1.

We further confirmed the translational repression activity of 22-nt siRNAs using an in vitro cell-free system<sup>17</sup>. A synthesized 22-nt or 21-nt siRNA targeted to the same site of *NIA2* transcript was assembled with AGO1 to form RISC (Extended Data Fig. 6a-d). After incubating the RISC with *NIA2* transcripts, 21-nt siRNAs showed more rapid cleavage than 22-nt siRNAs (Fig. 3d, e), partly because of the higher AGO1-loading efficiency of 21-nt siRNA (Extended Data Fig. 6d). Meanwhile, 22-nt and



**Fig. 3 | 22-nt siRNAs mediate translational repression and siRNA amplification in an AGO1-dependent manner.** **a**, **b**, Images of 20-day-old plants of the indicated genotypes. Scale bars, 2 cm. **c**, IGV of AGO1-associated 22-nt siRNA abundance at *NIA1/2*. IP, immunoprecipitation. **d**, Northern blotting (NB) of *NIA2* mRNAs labelled with 3 $\times$  FLAG tag, and anti-FLAG western blotting (WB) of the 3 $\times$  FLAG *NIA2* protein at the indicated time points in the presence ('21-nt' or '22-nt') or absence ('No siRNA') of AGO1-RISC programmed with 21- or 22-nt *NIA2* siRNA in vitro. 18S ribosomal RNA and Ponceau S staining were used, respectively, as RNA and protein loading controls. Numbers to the left of the SDS-PAGE gels indicate the molecular mass (in kDa) of proteins. **e**, **f**, Relative accumulation of full-length 3 $\times$  FLAG *NIA2* mRNA (**e**) and 3 $\times$  FLAG

*NIA2* protein (**f**) from **d**. Data shown as relative amounts after normalization to the mRNA level at 1 min (**e**) or the protein level at 120 min (**f**) in the 'no siRNA' reaction (set as '1'). **g**, Ratio of relative 3 $\times$  FLAG *NIA2* protein and mRNA levels at indicated time points, using the quantitative results from **e**, **f**. Values from 1 min, 5 min and 15 min were not included because the protein level was undetectable at these time points. **h**, IGV of 22-nt siRNAs derived from *NIA1/2* genes. **i**, **j**, Heat map showing abundance of 22-nt siRNAs from 1,182 and 182 genes producing 22-nt siRNAs. The numbers of individual biological experiments are as follows: **a, b**,  $n = 3$ ; **c, d**,  $n = 2$ ; **h, j**,  $n = 4$ . For gel source data, see Supplementary Fig. 1. For sequences of 3 $\times$  FLAG *NIA2* and AGO1, see Supplementary Fig. 2.



**Fig. 4 | Functional analysis and environmental induction of 22-nt siRNAs.**

**a**, Images of eight-day-old seedlings of the indicated genotypes grown on Murashige and Skoog (MS) medium. Scale bar, 0.5 cm. **b**, Propidium iodide staining of 9-day-old plant roots of the indicated genotypes. The white triangles point to the upper boundaries of the root apical meristem region. Scale bar, 50  $\mu$ m. **c**, Abundance of 21- and 22-nt siRNAs from all genes, excluding miRNAs and tasiRNAs, in 10-day-old Col-0 and *dcl4-2* plants, grown on normal MS medium or nitrogen-depleted MS medium (MS-N). **d**, Accumulation of *NIA1/2*-generated 22-nt siRNAs in plants. **e**, Images of 12-day-old seedlings grown on indicated media. Scale bar, 5 mm. **f**, Relative *DCL2*, *DCL4*, *SKI2* and *EIN5* gene-expression levels in plants. NA, not applicable. **g**, *NIA1* and *NIA2* protein levels in plants. HSP90 was used as a control to normalize the protein level. *NIA1* and *NR* ratios represent the relative protein level under each

condition after normalization to *NIA1* or *NR* protein amount in Col-0 plants on MS medium (set as '1'). **h**, Relative *NIA1/2* gene expression levels in plants. **i, j**, Polysome profiling showing the global translation efficiency of 10-day-old *Arabidopsis* plants under nitrogen deficiency condition (**i**) or in *nia1 nia2* mutant (**j**). The sucrose gradient absorbance ( $A_{260}$ ) of fractions was monitored. **k**, Proposed model for translational repression by 22-nt siRNAs and its role in plant stress adaptation. The 22-nt siRNAs accumulate in response to environmental stress, such as nitrogen starvation, through the inhibition of RNA decay and *DCL4* activities. They further cause translational repression and trigger transitive sRNA amplification. Data are presented as means  $\pm$  s.d. (**c**, **f**, **h**). The numbers of individual biological experiments are as follows: **a**, **c**, **d**, **f**–**j**,  $n = 3$ ; **b**,  $n = 2$ ; **e**,  $n = 6$ . All *P*-values are from two-tailed Student's *t*-tests. For gel source data, see Supplementary Fig. 1.

21-nt siRNAs led to a marked decrease in *NIA2* protein levels (Fig. 3d, f). By calculating the ratio of *NIA2* protein levels to transcript levels, we found that 22-nt siRNAs tended to repress *NIA2* protein synthesis more efficiently than 21-nt siRNAs (Fig. 3g), supporting their potent ability to repress translation.

The finding that 22-nt miRNAs initiate phased tasiRNA production<sup>18,19</sup> prompted us to examine the function of complexes containing 22-nt siRNAs and AGO1 in triggering transitive sRNA amplification. Notably, 22-nt siRNAs—including those derived from *NIA1/2* and *SMXL4/5*—were greatly eliminated by *ago1-27* or another hypomorphic allele, *ago1-45* (ref. <sup>20</sup>), which similarly rescued the defects of *ein5 dcl4* and *ski2 dcl4* mutants (Fig. 3h–j and Extended Data Fig. 7a–d). Considering the modest effect of *ago1-45* and *ago1-27* mutations on sRNA loading and stability<sup>13,20</sup>, these results indicate that AGO1 function—particularly its ability to repress translation—is associated with efficient biogenesis of 22-nt siRNAs.

## 22-nt siRNAs in stress adaptation

We next investigated the biological roles of 22-nt siRNAs in regulating plant growth and adaptation. We found that *ski2 dcl4* mutants showed obvious root-growth inhibition and seedling retardation on soil, partially resembling *nia1 nia2* double mutants<sup>21</sup> (Fig. 4a, b and Extended Data Fig. 8a, b). Both *ein5 dcl4* and *ski2 dcl4* plants showed abnormal phloem development and downregulation of multiple phloem-related genes (*CALS7* and *NEN4*) (Extended Data Fig. 8c, d), in line with previous findings that *SMXL4/5* produced 22-nt siRNAs occasionally in *dcl4*

mutants<sup>22</sup>, and that phloem formation is impaired in *smxl4 smxl5* double mutants<sup>11</sup>. Excessive anthocyanin accumulation was observed in *ein5 dcl4* and *ski2 dcl4* plants (Extended Data Fig. 8e), consistent with previous reports that disturbed nitrogen metabolism or defective phloem formation activates anthocyanin biogenesis<sup>11,23</sup>. All of these growth defects and the altered expression of phloem-related genes observed in *ein5 dcl4* and *ski2 dcl4* can be rescued by *dcl2* (Fig. 4a, b and Extended Data Fig. 8). Furthermore, gene-ontology analysis of differentially expressed genes revealed that myriad plant stress-response pathways (including, for example, pathways involving the hormone abscisic acid (ABA) or the response to salt) were activated, whereas plant growth responses (for example, photosynthesis and cell division) were inhibited in *ein5 dcl4* and *ski2 dcl4* plants (Extended Data Fig. 9a–j and Supplementary Tables 3, 4). Together, these results suggest an important role of 22-nt siRNAs in fine-tuning plant growth and stress responses.

We next explored the functionality of endogenous 22-nt siRNAs under environmental stress conditions. In wild-type plants, the production of 22-nt siRNAs was noticeably induced when nitrogen was deficient; this induction was more evident in *dcl4* mutants (Fig. 4c)—for example, in the case of *NIA1/2*-derived 22-nt siRNAs (Fig. 4d). Accordingly, wild-type plants displayed shortened root length, altered leaf pigmentation and decreased *NIA1/2* protein abundance upon nitrogen deficiency, while *dcl4* mutants showed hypersensitive responses (Fig. 4e). One possible explanation could be the induction of *DCL2* and downregulation of RNA-decay genes (*SKI2* and *EIN5*) and *DCL4* (Fig. 4f). It has also previously been reported that *DCL4* activity is compromised when nitrogen is deficient<sup>24</sup>. Notably, nitrogen depletion led to an almost complete



loss of NIA1/2 proteins in wild-type plants (Fig. 4g), which can hardly be explained by the moderate accumulation of 22-nt siRNAs alone. Given that the *NIA1/2* genes are nitrate-induced<sup>25</sup>, a substantial reduction of *NIA1/2* expression upon nitrogen depletion might also be involved (Fig. 4h). Moreover, polysome profiling revealed that loss of NIA1/2 proteins—in either nitrogen-depleted plants or *nia1 nia2* mutants—resulted in a reduction of ribosome-associated mRNAs, particularly in fractions 1–5 (Fig. 4i, j). This global translational slowdown is probably due to a shortage of amino acids, the synthesis of which relies highly on NIA1/2- and nitrite-reductase-dependent ammonium production<sup>26</sup>. Thus, we propose that under conditions of nitrogen starvation, in addition to reduced *NIA1/2* expression, compromised RNA decay and DCL4 activity cause an increase in 22-nt siRNAs and translational repression of *NIA1/2* mRNAs. This combinatory effect causes a substantial loss of NIA1/2 proteins and thus reduces amino-acid levels, slowing global translation and restraining plant growth. This is reminiscent of the response of yeast to nutrient deprivation, in which translation is inhibited by decreasing the splicing of genes that encode ribosomal proteins in order to enhance survival<sup>27,28</sup>.

Meanwhile, we found that treatment with ABA also caused notable accumulation of 22-nt siRNAs in *dcl4* plants, with 90% of them being derived from *NIA1/2*, contributing to the translational repression of *NIA1/2* (Extended Data Fig. 10a–e). Accordingly, the *dcl4* mutant was hypersensitive to ABA treatment and displayed more severe growth inhibition than wild-type plants (Extended Data Fig. 10f). By searching and reanalysing previously published databases, we found that high salinity also led to an elevation of overall 22-nt siRNA production in wild-type plants, including siRNAs derived from *NIA1/2*. As expected, salt treatment decreased the accumulation of NIA1/2 proteins (Extended Data Fig. 10g–i). Taken together, these data prompt us to propose that the production of 22-nt siRNAs, particularly from *NIA1/2*, might be essential for plants to adapt to environmental stresses. The efficient repression of NIA1 and NIA2 could be a common regulatory mechanism when plants encounter environmental stresses that allows them to deliberately switch from growth to defence.

## Discussion

Here we have reported the massive production of endogenous 22-nt siRNAs, particularly from the *NIA1/2* genes, in plants deficient in both DCL4 activity and cytoplasmic RNA decay. This increased biogenesis of 22-nt siRNAs requires RDR6 and DCL2 activity. AGO1 is required for both their biosynthesis and their functionality. Distinct from 21-nt and 24-nt siRNAs, 22-nt siRNAs are efficient in mediating translational repression and less effective in target cleavage, and also trigger transitive RNA interference and silencing amplification (Fig. 4k). The induction of 22-nt siRNAs under certain stress conditions leads to translational slowdown and growth inhibition. At present, it is unknown how 22-nt siRNAs mediate translational repression and transitive RNA interference. It is conceivable that AGO1 loaded with 22-nt siRNAs may prefer to stall ribosomes—a phenomenon observed for some 22-nt miRNAs<sup>13,29</sup>. A 22-nt-siRNA-AGO1 complex associated with stalled ribosomes or cleaved mRNA fragments would recruit RDR6 to initiate dsRNA synthesis and siRNA amplification.

## Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information,

acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-020-2231-y>.

- Guo, Z., Li, Y. & Ding, S. W. Small RNA-based antimicrobial immunity. *Nat. Rev. Immunol.* **19**, 31–44 (2019).
- Borges, F. & Martienssen, R. A. The expanding world of small RNAs in plants. *Nat. Rev. Mol. Cell Biol.* **16**, 727–741 (2015).
- Bologna, N. G. & Voinnet, O. The diversity, biogenesis, and activities of endogenous silencing small RNAs in *Arabidopsis*. *Annu. Rev. Plant Biol.* **65**, 473–503 (2014).
- Song, X., Li, Y., Cao, X. & Qi, Y. MicroRNAs and their regulatory roles in plant-environment interactions. *Annu. Rev. Plant Biol.* **70**, 489–525 (2019).
- Henderson, I. R. et al. Dissecting *Arabidopsis thaliana* DICER function in small RNA processing, gene silencing and DNA methylation patterning. *Nat. Genet.* **38**, 721–725 (2006).
- Xie, Z., Allen, E., Wilken, A. & Carrington, J. C. DICER-LIKE 4 functions in trans-acting small interfering RNA biogenesis and vegetative phase change in *Arabidopsis thaliana*. *Proc. Natl Acad. Sci. USA* **102**, 12984–12989 (2005).
- Deleris, A. et al. Hierarchical action and inhibition of plant Dicer-like proteins in antiviral defense. *Science* **313**, 68–71 (2006).
- Zhang, X. et al. Suppression of endogenous gene silencing by bidirectional cytoplasmic RNA decay in *Arabidopsis*. *Science* **348**, 120–123 (2015).
- Bouché, N., Lauresergues, D., Gascolli, V. & Vaucheret, H. An antagonistic function for *Arabidopsis* DCL2 in development and a new function for DCL4 in generating viral siRNAs. *EMBO J.* **25**, 3347–3356 (2006).
- Wilkinson, J. Q. & Crawford, N. M. Identification and characterization of a chlorate-resistant mutant of *Arabidopsis thaliana* with mutations in both nitrate reductase structural genes *NIA1* and *NIA2*. *Mol. Gen. Genet.* **239**, 289–297 (1993).
- Wallner, E. S. et al. Strigolactone- and karrikin-independent SMXL proteins are central regulators of phloem formation. *Curr. Biol.* **27**, 1241–1247 (2017).
- Brodersen, P. et al. Widespread translational inhibition by plant miRNAs and siRNAs. *Science* **320**, 1185–1190 (2008).
- Li, S. et al. Biogenesis of phased siRNAs on membrane-bound polysomes in *Arabidopsis*. *eLife* **5**, e22750 (2016).
- Yu, B. et al. Methylation as a crucial step in plant microRNA biogenesis. *Science* **307**, 932–935 (2005).
- Li, S. et al. MicroRNAs inhibit the translation of target mRNAs on the endoplasmic reticulum in *Arabidopsis*. *Cell* **153**, 562–574 (2013).
- Mi, S. et al. Sorting of small RNAs into *Arabidopsis* argonaute complexes is directed by the 5' terminal nucleotide. *Cell* **133**, 116–127 (2008).
- Tomari, Y. & Iwakawa, H. O. In vitro analysis of ARGONAUTE-mediated target cleavage and translational repression in plants. *Methods Mol. Biol.* **1640**, 55–71 (2017).
- Cuperus, J. T. et al. Unique functionality of 22-nt miRNAs in triggering RDR6-dependent siRNA biogenesis from target transcripts in *Arabidopsis*. *Nat. Struct. Mol. Biol.* **17**, 997–1003 (2010).
- Chen, H. M. et al. 22-Nucleotide RNAs trigger secondary siRNA biogenesis in plants. *Proc. Natl Acad. Sci. USA* **107**, 15269–15274 (2010).
- Smith, M. R. et al. Cyclophilin 40 is required for microRNA activity in *Arabidopsis*. *Proc. Natl Acad. Sci. USA* **106**, 5424–5429 (2009).
- Wang, R. et al. Genomic analysis of the nitrate response using a nitrate reductase-null mutant of *Arabidopsis*. *Plant Physiol.* **136**, 2512–2522 (2004).
- Wu, Y. Y. et al. DCL2- and RDR6-dependent transitive silencing of SMXL4 and SMXL5 in *Arabidopsis dcl4* mutants causes defective phloem transport and carbohydrate over-accumulation. *Plant J.* **90**, 1064–1078 (2017).
- Rubin, G., Tohge, T., Matsuda, F., Saito, K. & Scheible, W. R. Members of the LBD family of transcription factors repress anthocyanin synthesis and affect additional nitrogen responses in *Arabidopsis*. *Plant Cell* **21**, 3567–3584 (2009).
- Seta, A. et al. Post-translational regulation of the dicing activities of *Arabidopsis* DICER-LIKE 3 and 4 by inorganic phosphate and the redox state. *Plant Cell Physiol.* **58**, 485–495 (2017).
- Wang, R. et al. Multiple regulatory elements in the *Arabidopsis NIA1* promoter act synergistically to form a nitrate enhancer. *Plant Physiol.* **154**, 423–432 (2010).
- Campbell, W. H. & Kinghorn, K. R. Functional domains of assimilatory nitrate reductases and nitrite reductases. *Trends Biochem. Sci.* **15**, 315–319 (1990).
- Parenteau, J. et al. Introns are mediators of cell response to starvation. *Nature* **565**, 612–617 (2019).
- Morgan, J. T., Fink, G. R. & Bartel, D. P. Excised linear introns regulate growth in yeast. *Nature* **565**, 606–611 (2019).
- Hou, C. Y. et al. Global analysis of truncated RNA ends reveals new insights into ribosome stalling in plants. *Plant Cell* **28**, 2398–2416 (2016).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2020



## Methods

No statistical methods were used to determine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

### Plant materials and growth conditions

All mutants used in this study were in the Col-0 background. *Arabidopsis thaliana* mutants, including *ein5-1* (ref. <sup>30</sup>), *ski2-2* and *ski2-3* (ref. <sup>8</sup>), *dcl2-1* (ref. <sup>31</sup>), *dcl4-2* (ref. <sup>5</sup>), *rdm6-11* (ref. <sup>32</sup>), *ago1-45* (ref. <sup>20</sup>), *ago1-27* (ref. <sup>33</sup>), *hen1-8* (ref. <sup>34</sup>), *ein5-1 dcl4-2*, *ski2-2 dcl4-2*, *ein5-1 ski2-3*, *ein5-1 dcl4-2 dcl2-1*, *ski2-2 dcl4-2 dcl2-1*, and *ein5-1 ski2-3 dcl4-2 dcl2-1* (ref. <sup>8</sup>) were described previously. The *nial1-3* (Salk 148487C) and *nial2-1* (Salk 138297C) mutants were obtained from the *Arabidopsis* Biological Resource Center (<https://abrc.osu.edu/>). Other double and triple mutants used here were generated by genetic cross.

We used MS basal salt and nitrogen-depleted MS basal salt from Phyto Technology (catalogue numbers M524 and M531) to prepare full-nutrition MS medium and nitrogen-depleted MS (MS-N) medium. Surface-sterilized seeds were plated on the media (pH 5.7) for 3 days at 4 °C and then kept in the greenhouse for another 6–7 days (22 °C; 16 h/8 h photoperiod) before being transferred to soil.

### RNA-seq and sRNA-seq

Total RNA and sRNA were extracted from the aerial part of 20-day-old plants (6 days on MS medium and 14 days on soil) or 10-day-old seedlings on MS or MS-N medium. For mRNA sequencing (mRNA-seq), we extracted total RNA using the RNeasy Plant Mini Kit (Qiagen). Libraries were prepared using an mRNA preparation kit (Illumina), following the standard protocol. Sequencing of paired-end 150-base-pair mRNAs was conducted by BGI (Shenzhen) company using the Illumina HiSeqX ten platform. Raw reads were filtered by Trimmomatic (<http://www.usadellab.org/cms/?page=trimmomatic>)<sup>35</sup> with default parameters. Filtered reads were then aligned to the *Arabidopsis* Col-0 genome (TAIR10 release) using the tophat2 program<sup>36</sup> with two mismatches allowed. Differentially expressed genes were called by cuffdiff<sup>37</sup>.

For sRNA sequencing (sRNA-seq), sRNA was extracted using the miRNeasy Mini Kit (Qiagen). We prepared sRNA libraries using the NEBNext Small RNA Library Prep Set for Illumina (catalogue number E7330L). Single-end 50-bp sRNA sequencing was conducted by BGI (Shenzhen) using the Illumina HiSeq 4000 platform. Raw reads were filtered using the fastx\_toolkit ([http://hannonlab.cshl.edu/fastx\\_toolkit/](http://hannonlab.cshl.edu/fastx_toolkit/)). The 18–26-nt clean reads were mapped to a structural RNA (ribosomal RNAs, transfer RNAs, small nucleolar RNAs and small nuclear RNAs) database (<http://rfam.xfam.org/>). Then the unmapped reads were aligned to the TAIR10 genome using bowtie<sup>38</sup> with following parameters: -v 0 -a -M 3 -best -strata. Differential small RNA-generating genes were called using the DESeq package in R (<https://bioconductor.org/packages/release/bioc/html/DESeq.html>). Gene-ontology enrichment analysis was performed using DAVID Bioinformatics Resources 6.8 (ref. <sup>39</sup>) and the BiNGO package in Cytoscape software<sup>40</sup> with default parameters. Mapped reads were visualized using IGV<sup>41</sup>, version IGV\_2.3.72.

Genes generating 22-nt siRNAs in *ein5-1 dcl4-2* (*ed*) or *ski2-2 dcl4-2* (*sd*) plants were identified using the following criteria: (i) 22-nt siRNA reads in *ed* or *sd* plants should be at least twofold higher than those in Col-0, *dcl4-2* and corresponding RNA-decay mutants (*ein5-1* for *ed* or *ski2-2* for *sd*); (ii) 22-nt siRNA reads in *ed* or *sd* plants should be at least twofold higher than their corresponding triple mutants generated by crossing with *dcl2-1* (*ein5-1 dcl4-2 dcl2-1* or *ski2-2 dcl4-2 dcl2-1*); (iii) the false discovery rate (FDR) of all comparisons should be less than 0.01; (iv) the sum of 22-nt siRNA reads in seven libraries (Col-0, *ein5-1*, *dcl4-2*, *dcl2-1*, *ein5-1 dcl4-2*, *ein5-1 dcl2-1* and *ein5-1 dcl4-2 dcl2-1* for *ein5-1 dcl4-2*; Col-0, *ski2-2*, *dcl4-2*, *dcl2-1*, *ski2-2 dcl4-2*, *ski2-2 dcl2-1* and *ski2-2 dcl4-2 dcl2-1* for *ski2-2 dcl4-2*) should be larger than 21.

To calculate the abundance of 21- or 22-nt siRNAs generated from all genes except for miRNAs and tasiRNAs, we added up 21- or 22-nt siRNAs reads from the 5'-UTR, exonic regions, intronic regions and 3'-UTR and normalized them to total mapped 18–26-nt siRNAs reads. To calculate 21- or 22-nt siRNAs generated from *TAS* genes, we added up 21- or 22-nt siRNA reads from all eight annotated *TAS* genes and normalized them to total mapped 18–26-nt siRNAs reads.

### sRNA phasing score analysis

Phasing analysis was performed as described<sup>42</sup> using ShortStack<sup>43</sup> with default parameters (version 3.8.5).

### Polysome profiling

Polysome profiling was performed as described<sup>44</sup> with minor modifications. In brief, the aerial parts of 20-day-old (Fig. 2f and Extended Data Fig. 3d) or 10-day-old (Fig. 4i, j) plants were collected and ground into fine powder in liquid nitrogen, then homogenized in 1 ml of pre-cold polysome extraction buffer (PEB: 100 mM Tris-HCl, pH 9.0, 100 mM KCl, 20 mM MgCl<sub>2</sub>, 12.5 mM EGTA, pH 8.0; 75 mM NaCl, 10 mM Tris-HCl, pH 7.4, 3 mM dithiothreitol (DTT), 100 mM sucrose, 37.5 µg ml<sup>-1</sup> cycloheximide, 25 µg ml<sup>-1</sup> chloramphenicol, 6.25 µl 20% detergent mix (20% Brij-35, 20% Triton X-100, 20% Igopal CA630, 20% Tween 20); 25 µl Triton X-100). After thawing on ice for 10 min, the lysates were centrifuged at 4 °C, 14,500 r.p.m. for 20 min. The supernatants were transferred to a 40-µm filter falcon tube and centrifuged at 4 °C, 14,500 r.p.m. for another 15 min. The supernatants were then transferred into a 1.5-ml tube. For each sample, 700 µl of lysate was loaded on top of a 10–50% sucrose gradient, and another 70 µl of lysate was used as input for total mRNA isolation using Trizol reagent. The samples were then centrifuged at 4 °C, 37,000 r.p.m. for 2 h and 10 min in a P40ST rotor. Polysomal RNAs distributed along the sucrose gradient according to their different sedimentation coefficient after ultracentrifugation, and were separated using a gradient fractionator system (BioComp Instruments). Fractions containing monosomes to the heaviest polysomes were collected and subjected to RNA isolation using Trizol followed by quantitative real-time polymerase chain reaction (PCR). Polysome distributions of *NIA1/2* (Fig. 2e and Extended Data Fig. 3e–g) and *GTE2/7* (Extended Data Fig. 2g) mRNA levels were normalized against *PP2AA3* (*AT1G13320*). *NIA1/2* and *GTE2/7* expression in each polysomal fraction was calculated as the percentage of its expression in total RNA.

### Total protein extraction and immunoblot assays

Plant materials were ground into fine powder in liquid nitrogen and then homogenized in protein extraction buffer (100 mM Tris-HCl, pH 6.8, 4% SDS, 10% glycerol, 50 mM DTT). The samples were heated for 10 min at 65 °C and then centrifuged for 10 min at 13,000 r.p.m. at 4 °C. After centrifugation, the supernatants were separated on 7.5% SDS-PAGE gels, and the proteins were transferred to polyvinylidene difluoride (PVDF) membranes (Millipore). Immunoblot assays were performed to analyse the samples using specific antibodies, including anti-AGO1 antibody (Agrisera, catalogue number AS09527), anti-HSP90 antibody (Beijing Protein Innovation, catalogue number AbM51099-31-PU) and anti-NR antibody (Agrisera, catalogue number AS08310). Goat anti-mouse (or anti-rabbit) immunoglobulin G (IgG) (H+L) horseradish peroxidase (HRP)-conjugated secondary antibody (Promega Corporation) was diluted 10,000-fold. Anti-NIA1 specific antibody was made by Abclonal Biotechnology using the peptide KIVKETEVITTVDSY DD. Western blot images were captured using a Tanon 5200 Multi chemiluminescent imaging system (Tanon).

### Gene-expression analysis by quantitative PCR

Two micrograms of total RNA extracted using Trizol (Invitrogen) were used for reverse transcription at 42 °C for 1 h with Moloney murine leukaemia virus (M-MLV) reverse transcriptase (Promega) according to the manufacturer's instructions. After tenfold dilution

of complementary DNA product, real-time PCR was performed using SYBR Premix Ex Taq II (Tli RNaseH Plus) (Takara Bio) with a Roche LightCycler 96 platform. *PP2A43* (AT1G13320) was used as a reference gene. We used 40 cycles of amplification for all genes tested herein, except for *SMXL4* and *SMXL5*, which require 50 cycles owing to their low expression levels. Gene-specific primers used here are listed in Supplementary Table S.

### Cytoplasmic-nuclear fractionation

*Arabidopsis* cytoplasmic-nuclear fractionation was performed as described<sup>45,46</sup>. In brief, 20-day-old plants were collected and ground into a fine powder in liquid nitrogen and then homogenized in 1 ml lysis buffer (20 mM Tris-HCl, pH 7.5, 20 mM KCl, 2 mM EDTA, 2.5 mM MgCl<sub>2</sub>, 25% glycerol, 250 mM sucrose, 5 mM DTT, 1,000 U ml<sup>-1</sup> RNase inhibitor; 2× protease-inhibitor cocktail, 1 mM phenylmethanesulfonyl fluoride (PMSF)) by gentle shaking. The homogenate was filtered through a double-layer miracloth. The filtered homogenate was centrifuged at 4 °C at 1,500g for 10 min to pellet the nuclei, and the supernatant consisting of the cytoplasmic fraction was centrifuged at 4 °C at 15,000g for another 10 min to further remove debris. Then 800 µl of supernatant was collected for downstream RNA extraction to perform reverse transcription and quantitative reverse transcription (qRT)-PCR analysis. The nuclei pellet was washed six times with 1 ml NRBT buffer (20 mM Tris-HCl, pH 7.4, 25% glycerol, 2.5 mM MgCl<sub>2</sub>, 0.2% Triton X-100 and 20 U ml<sup>-1</sup> RNase inhibitor). The sample was centrifuged at 4 °C, 1,500g, for 10 min. The supernatant was discarded and the nuclei resuspended with 200 µl nuclei storage buffer (NSB: 20 mM Tris-HCl, pH 7.5, 25% glycerol, 2.5 mM MgCl<sub>2</sub>, 440 mM sucrose, and 1× protease-inhibitor cocktail). For quality control of cytoplasmic-nuclear fractionation, we used 100 µl of the total and cytoplasmic fractions and 30 µl of the nuclear fraction for protein extraction and immunoblot assay. We used tubulin protein (antibody from Abclonal, catalogue number AC010) as a cytoplasmic marker and histone H3 (antibody from Abclonal, A2348) as a nuclear marker.

### Preparation of 3×FLAG-NIA2 mRNA

We constructed pBYL-3×FLAG-NIA2 by inserting the coding sequence of the *NIA2* gene with an amino-terminal 3×FLAG tag into the *Ascl* site of pBYL2 (ref. <sup>47</sup>). The 3×FLAG-NIA2 mRNA was in vitro transcribed from *NotI*-linearized pBYL-3×FLAG-NIA2 using an AmpliScribe T7 high yield transcription kit (Lucigen), followed by capping with a ScriptCap m<sup>7</sup>G capping system (Cell Script) and poly(A)-tailing with A-Plus poly(A) polymerase tailing kit (Cell Script).

### In vitro RNA silencing assay

In vitro RNA silencing was essentially performed as described<sup>17</sup>, with slight modifications. In brief, 15.6 µl of BY-2 lysate, 7.8 µl of substrate mixture (containing ATP, ATP-regeneration system, and amino-acid mixture), and 1.56 µl of 300 nM AGO1 mRNA were mixed, and incubated at 25 °C for 30 min. Then, 3.12 µl of 1.5 µM NIA2 siRNA duplexes (GeneDesign) were added to the reaction, and incubated at 25 °C for 90 min. Subsequently, 1.5 µl of 20 nM 3×FLAG-NIA2 mRNA was added. At each time point, 2-µl aliquots of the reaction mixture were transferred into two new tubes, and used for western and northern blots. For detection of 3×FLAG-NIA2, anti-DDDDK-tag monoclonal antibody (diluted at 1/5,000) (Medical and Biological Laboratories, catalogue number M185-3L) was used as the primary antibody. Anti-IgG (H+L) (mouse) pAb-HRP (diluted at 1/5,000) (Medical and Biological Laboratories) was used as secondary antibody. The 5'-<sup>32</sup>P-radiolabelled DNA oligonucleotide mixture (5'-acctggctcgagacgagcgattggcgattactacagaggccgcg-3', 5'-ccatggcatattccttcttgatgctcggttcctgattggaccag-3', 5'-ccttcttcgctcgcgagattcggcgtctaccaaagaaggttaaaact-3', 5'-taatttggttaaacggaggttattccggttcacactgccaacattg-3') was used for detection of 3×FLAG-NIA2 mRNA.

### In vitro RISC assembly assay

The 3×FLAG-AGO1 mRNA<sup>48</sup> was in vitro translated as described above. Then, the radiolabelled NIA2-siRNA duplex was added to the reaction mixture, and incubated at 25 °C for 90 min. Subsequently, the reaction mixture was incubated with Dynabeads protein G coated with anti-Flag M2 antibody (Sigma-Aldrich) on a rotator at 4 °C for 1 h. Then, the beads were washed three times with lysis buffer (30 mM HEPES, pH 7.4, 100 mM KOAc and 2 mM Mg(OAc)<sub>2</sub>) containing 1% TritonX-100 and 800 mM NaCl. After treatment with proteinase K, the samples were mixed with an equal volume of 2× formamide dye<sup>17</sup>, incubated for 5 min at 68 °C, and loaded onto a 15% urea PAGE denaturing gel. After drying, the gel was analysed using PhosphorImager (Typhoon FLA 7000, GE Healthcare). It took 15–30 min for the in vitro system to produce detectable translation products once the mRNA was added; hence, no band was detected in the first three lanes of each condition in the western blot section (Fig. 3d).

### Anthocyanin content measurement

The anthocyanin content was measured as reported<sup>49</sup>. Samples used for anthocyanin extraction were preweighted (usually 0.1–0.5 g), and 1 ml anthocyanin extraction buffer (propanol/HCl/H<sub>2</sub>O at 18/1/81) was added to each sample. After overnight incubation at room temperature in the dark, samples were centrifuged at 14,000 r.p.m. for 20 min. Supernatants were diluted twofold before measuring the absorbance (A) at 535 nm and 650 nm, and then the relative anthocyanin content was calculated using the following formula:  $2 \times (A_{535} - A_{650})$  per gram of fresh weight.

### Aniline staining of callose depositions in cotyledons

Aniline staining was performed as described<sup>50</sup>. In brief, *Arabidopsis* plants were grown on MS median for 6 days and then transferred to soil for another 12 days. Detached cotyledons were destained in acetic acid/ethanol (1/3) buffer over night until the cotyledons were transparent; the saturated destaining solution was replaced during the process. Then the destained cotyledons were washed three times with 150 mM K<sub>2</sub>HPO<sub>4</sub>, each time 10 min. After wash, the cotyledons were incubated in 150 mM K<sub>2</sub>HPO<sub>4</sub> buffer containing 0.01% aniline blue (staining solution) for 5 h in a Falcon tube wrapped in aluminium foil for light protection. After that, the samples were embedded in 50% glycerol for further observation. Callose deposition was observed using a fluorescence microscope (Zeiss Axio Imager M2), with an excitation wavelength of 370 nm for aniline blue. The emission maximum for aniline blue is 509 nm.

### Measurement of root length

For quantitative root-length measurements, images of 8-day-old seedlings grown on vertical MS medium were captured using a digital camera. Roots of 20 seedlings were measured using ImageJ software (<https://imagej.nih.gov/ij/>). Values shown are average lengths (means ± s.d.) of 20 roots of the indicated genotypes.

### AGO1 immunoprecipitation and sRNA-seq

Total protein was extracted from the aerial parts of 20-day-old plants with immunoprecipitation buffer (50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 10% glycerol; 0.1% NP-40, 5 mM MgCl<sub>2</sub>, 2× protease inhibitor cocktail). Samples were incubated with immunoprecipitation buffer at 4 °C for 40 min with gentle rotation before centrifugation. After a first centrifugation at 4 °C, 13,000 r.p.m., for 15 min, the supernatants were transferred to a new 2-ml tube and centrifuged at 4 °C, 13,000 r.p.m., for another 25 min. Then the extracts were incubated with AGO1 antibody for 2 h at 4 °C with rotation. Pierce protein A/G magnetic beads (Thermo Scientific) were added and incubated for another 2 h at 4 °C with rotation. Beads were collected with a DynaMag-2 magnet (Thermo Scientific) and washed six times with wash buffer (50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 10% glycerol, 0.5% NP-40; 5 mM MgCl<sub>2</sub>,

1× proteinase-inhibitor cocktail), each time for 5 min. Then sRNAs were extracted from immunoprecipitates using 800 µl Trizol reagent, precipitated with a twofold volume of ethanol, and subjected to sRNA library construction using a NEBNext small RNA library prep set for Illumina (NEB, catalogue number E7330L). Libraries were sequenced with the Illumina Novaseq 6000 platform under paired-end, 150-bp conditions; two biological replicates were performed for each genotype. Read 1 of each paired-end read was used for further analysis as in the sRNA-seq analysis pipeline. Sequence-conservation analysis of AGO1-associated sRNAs was performed using Weblogo 3 with default parameters (<http://weblogo.threeplusone.com>).

## Propidium iodide staining of *Arabidopsis* roots

Roots of 9-day-old *Arabidopsis* seedlings were stained with 1 µg ml<sup>-1</sup> propidium iodide solution (stock solution 1 mg ml<sup>-1</sup>; Thermo Scientific, P3566) for approximately 5 min on a slide. A coverslip was then placed on the top for confocal microscopy (Zeiss LSM880) with 40× water immersion. The excitation wavelength and emission maximum of propidium iodide are, respectively, 536 nm and 617 nm. We used 14 (Col-0 and *ski2-2*) or 15 (*dcl4-2*, *ski2-2 dcl4-2* and *ski2-2 dcl4-2 dcl2-1*) captured root images to calculate the number of meristematic cortex cells.

## Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

## Data availability

Sequencing data are available at the NCBI Gene Expression Omnibus (GEO; <https://www.ncbi.nlm.nih.gov/geo/>) under accession number GSE136164. Source gel data for immunoblots and radiograms (Figs. 2–4 and Extended Data Figs. 2, 4, 6, 10) are provided in Supplementary Fig. 1; source data for all graphs (Figs. 1–4 and Extended Data Figs. 1–3, 5, 7–10) are also provided and are available with the online version of the paper.

30. Olmedo, G. et al. ETHYLENE-INSENSITIVE5 encodes a 5'→3' exoribonuclease required for regulation of the EIN3-targeting F-box proteins EBF1/2. *Proc. Natl Acad. Sci. USA* **103**, 13286–13293 (2006).
31. Xie, Z. et al. Genetic and functional diversification of small RNA pathways in plants. *PLoS Biol.* **2**, E104 (2004).
32. Peragine, A., Yoshikawa, M., Wu, G., Albrecht, H. L. & Poethig, R. S. SGS3 and SGS2/SDE1/RDR6 are required for juvenile development and the production of trans-acting siRNAs in *Arabidopsis*. *Genes Dev.* **18**, 2368–2379 (2004).
33. Morel, J. B. et al. Fertile hypomorphic ARGONAUTE (*ago1*) mutants impaired in post-transcriptional gene silencing and virus resistance. *Plant Cell* **14**, 629–639 (2002).
34. Yu, B. et al. siRNAs compete with miRNAs for methylation by HEN1 in *Arabidopsis*. *Nucleic Acids Res.* **38**, 5844–5850 (2010).
35. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
36. Kim, D. et al. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14**, R36 (2013).

37. Trapnell, C. et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protocols* **7**, 562–578 (2012).
38. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
39. Huang, D., Sherman, B. T. & Lempicki, R. A. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* **37**, 1–13 (2009).
40. Shannon, P. et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).
41. Robinson, J. T. et al. Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–26 (2011).
42. Zhai, J. et al. MicroRNAs as master regulators of the plant NB-LRR defense gene family via the production of phased, trans-acting siRNAs. *Genes Dev.* **25**, 2540–2553 (2011).
43. Axtell, M. J. ShortStack: comprehensive annotation and quantification of small RNA genes. *RNA* **19**, 740–751 (2013).
44. Xu, G. et al. Global translational reprogramming is a fundamental layer of immune regulation in plants. *Nature* **545**, 487–490 (2017).
45. Wang, W. et al. An importin β protein negatively regulates microRNA activity in *Arabidopsis*. *Plant Cell* **23**, 3565–3576 (2011).
46. Xu, F. & Copeland, C. Nuclear extraction from *Arabidopsis thaliana*. *Bio Protoc.* **2**, e306 (2012).
47. Mine, A. et al. Identification and characterization of the 480-kilodalton template-specific RNA-dependent RNA polymerase complex of red clover necrotic mosaic virus. *J. Virol.* **84**, 6070–6081 (2010).
48. Endo, Y., Iwakawa, H. O. & Tomari, Y. *Arabidopsis* ARGONAUTE7 selects miR390 through multiple checkpoints during RISC assembly. *EMBO Rep.* **14**, 652–658 (2013).
49. Deikman, J. & Hammer, P. E. Induction of anthocyanin accumulation by cytokinins in *Arabidopsis thaliana*. *Plant Physiol.* **108**, 47–57 (1995).
50. Schenk, S. T. & Schikora, A. Staining of callose depositions in root and leaf tissues. *Bio Protoc.* **5**, e1429 (2015).

**Acknowledgements** We thank X. Chen for assistance with AGO1-immunoprecipitation experiments; J. Jia for help with sRNA phase analysis; K. Kiyokawa for assistance with plasmid constructions and mRNA preparation; and Y. Tomari, A. Hutchins and P. Pimpl for critical comments on the manuscript and language editing. This work was supported by the National Natural Science Foundation of China (grant 91740203 to H.G.), the National Key Research and Development Program of China (grant 2018YFA0507101 to H.G.), the Program for Guangdong Introducing Innovative and Entrepreneurial Teams (grants 2016ZT06S172 to J.Z. and B.L.), the Shenzhen Sci-Tech Fund (grant KYTDPT 20181011104005 to J.Z. and B.L.), Grants-in-Aid for Scientific Research on Innovative Areas ('Nascent-chain Biology'; grant 26116003 to H.-o.I.) and JST, PRESTO (grant JPMJPR 18K2 to H.-o.I.).

**Author contributions** H.G., H.W. and B.L. conceived the project and designed the experiments; H.W. and X.Z. prepared the genetic materials; H.W. collected genetic phenotypes and carried out western blot assays, qRT-PCR assays and AGO1-immunoprecipitation sRNA sequencing with contributions from X.T., X.X. and S.S.; B.L. and Y.P. carried out sRNA-seq and mRNA-seq bioinformatics analyses; L.F., H.Z. and J.Z. performed sRNA phase analysis; Y.P., Z.T. and Q.L.-h. conducted polysome profiling; Y.L. carried out cytoplasmic and nuclear fraction isolation assays; H.-o.I. performed in vitro RNA silencing assays; H.G., H.W. and B.L. wrote the manuscript with input from all authors.

**Competing interests** Authors declare no competing interests.

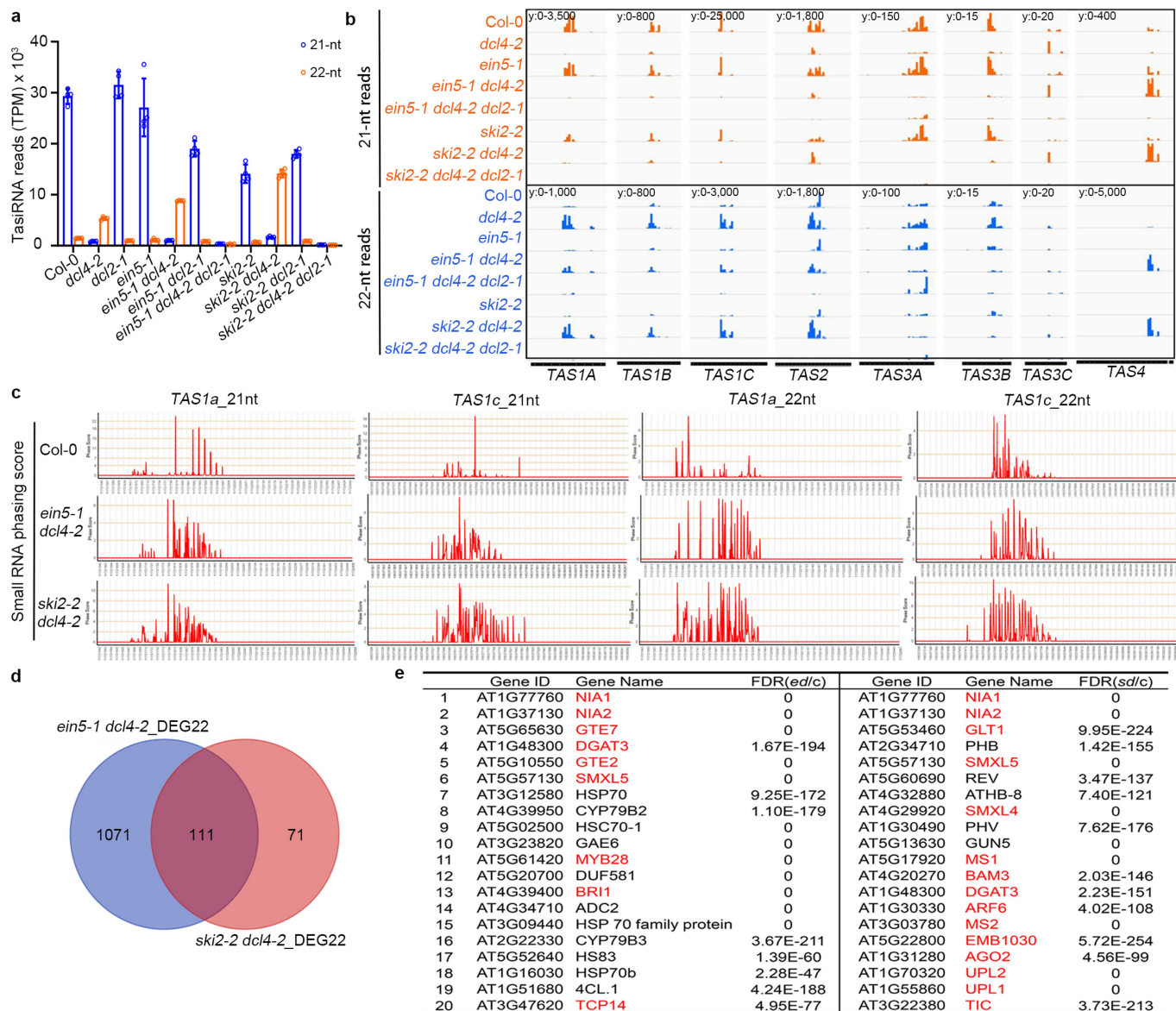
## Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41586-020-2231-y>.

**Correspondence and requests for materials** should be addressed to H.G.

**Peer review information** Nature thanks Peter Waterhouse and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

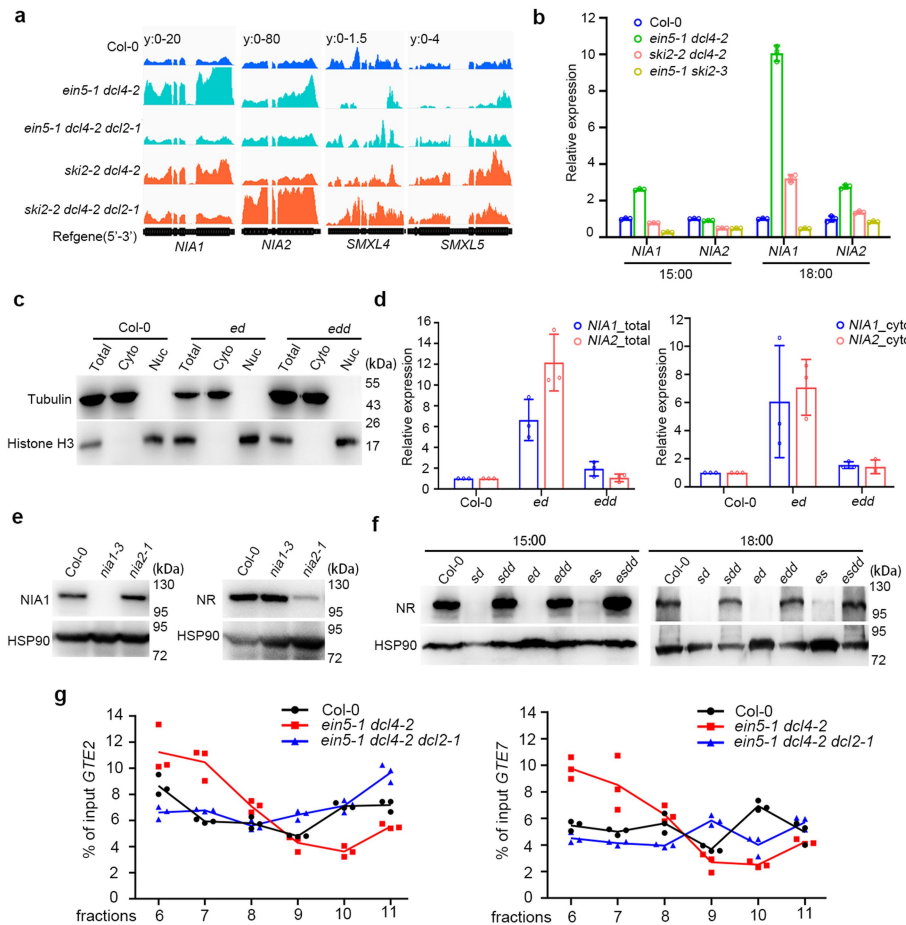
**Reprints and permissions information** is available at <http://www.nature.com/reprints>.



**Extended Data Fig.1 | Differential expression analysis of siRNAs in *ein5-1 dcl4-2* and *ski2-2 dcl4-2* plants.** **a**, Normalized tasiRNA abundance in each genotype of 20-day-old plants. TPM, tags per million. Data are presented as means  $\pm$  s.d.;  $n = 4$  biologically independent samples. **b**, IGV of 21- and 22-nt siRNA abundance at eight *TAS* gene loci. **c**, 21- and 22-nt siRNA phasing score of tasiRNAs from representative *TAS1a* and *TAS1c* loci in the indicated genotypes. The y-axis shows the siRNA phasing score. See Methods for calling of siRNA

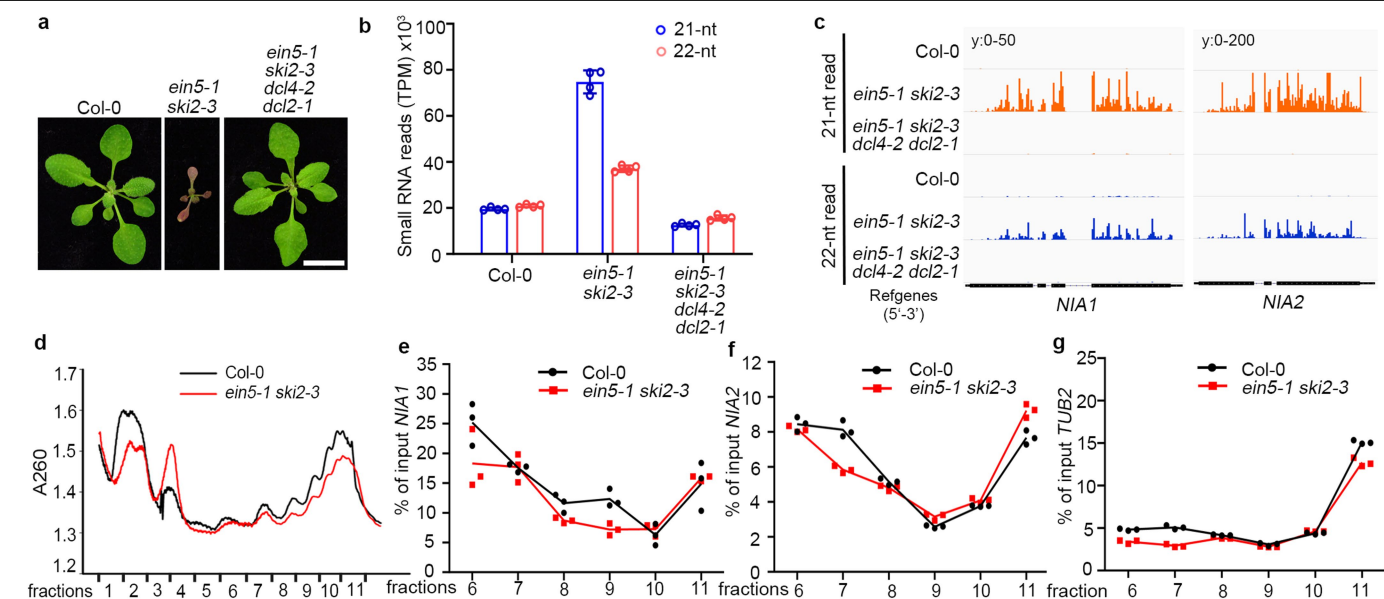
phasing scores. **d**, Venn diagram depicting 22-nt siRNA-generating genes that overlap between *ein5-1 dcl4-2* (1,182 genes) and *ski2-2 dcl4-2* (182 genes) plants. **e**, The top 20 gene loci and their FDRs in *ein5-1 dcl4-2* (left, *ed*) and *ski2-2 dcl4-2* (right, *sd*) plants, ranked by 22-nt siRNA abundance in each double mutant. Genes in red produce 22-nt siRNAs in both *ein5-1 dcl4-2* and *ski2-2 dcl4-2* plants. **c**, Col-0; **ed**, *ein5-1 dcl4-2*; **sd**, *ski2-2 dcl4-2*. FDR values are from Benjamini-Hochberg analysis;  $n = 4$  biologically independent samples.





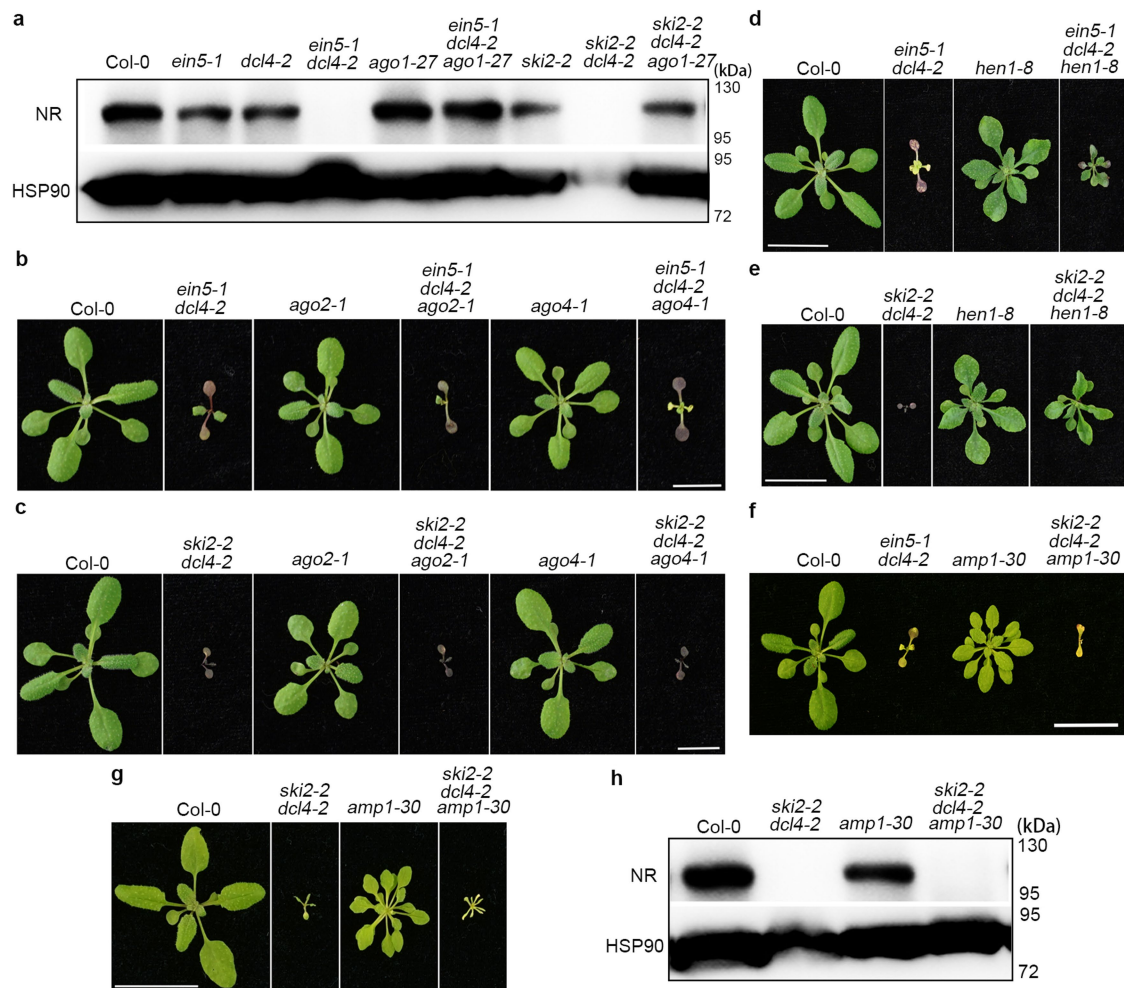
**Extended Data Fig. 2 | *NIA1* and *NIA2* mRNA and protein levels in *ein5-1 dcl4-2* and *ski2-2 dcl4-2* plants, and translational states of *GTE2/7* in *ein5-1 dcl4-2* plants. **a**, mRNA-seq read abundance of *NIA1/2* and *SMXL4/5* genes in the indicated genotypes. **b**, Relative expression levels of *NIA1/2* detected by qRT-PCR at 15:00 and 18:00 in 20-day-old Col-0, *ein5-1 dcl4-2*, *ski2-2 dcl4-2* and *ein5-1 ski2-3* plants. **c**, Separation of cytoplasmic and nuclear fractions of Col-0, *ein5-1 dcl4-2* (*ed*) and *ein5-1 dcl4-2 dcl2-1* (*edd*) plants. Tubulin and histone H3 proteins were used as cytoplasmic and nuclear markers, respectively. Cyto, cytoplasm; Nuc, nucleus. **d**, Relative expression level of *NIA1* and *NIA2* genes detected by qRT-PCR at 18:00 in the total and cytoplasmic fractions of 20-day-old plants. **e**, Verification of anti-NIA1 and anti-NR antibodies (the latter recognizing both**

*NIA1* and *NIA2*) using *nial1-3* and *nial2-1* null alleles. HSP90 was used as a loading control. **f**, Protein levels of *NIA1/2* detected by western blot at 15:00 and 18:00 in 20-day-old plants; *sd*, *ski2-2 dcl4-2*; *sdd*, *ski2-2 dcl4-2 dcl2-1*; *ed*, *ein5-1 dcl4-2*; *edd*, *ein5-1 dcl4-2 dcl2-1*; *es*, *ein5-1 ski2-3*; *esdd*, *ein5-1 ski2-3 dcl4-2 dcl2-1*. HSP90 was used as a loading control. **g**, *GTE2/7* mRNA levels were normalized against *PP2AA3* levels (locus AT1G13320). *GTE2/7* expression in each polysomal fraction was calculated as the percentage of its expression in total RNA. Data are shown as means  $\pm$  s.d. (**b**, **d**) or means with individual data points (technical replicates) (**g**). Numbers to the right of gels in **c**–**f** indicate the molecular mass (in kDa) of proteins. Numbers of individual biological experiments are: **a**,  $n = 4$ ; **b**–**d**,  $n = 3$ ; **e**–**g**,  $n = 2$ . For gel source data, see Supplementary Fig. 1.



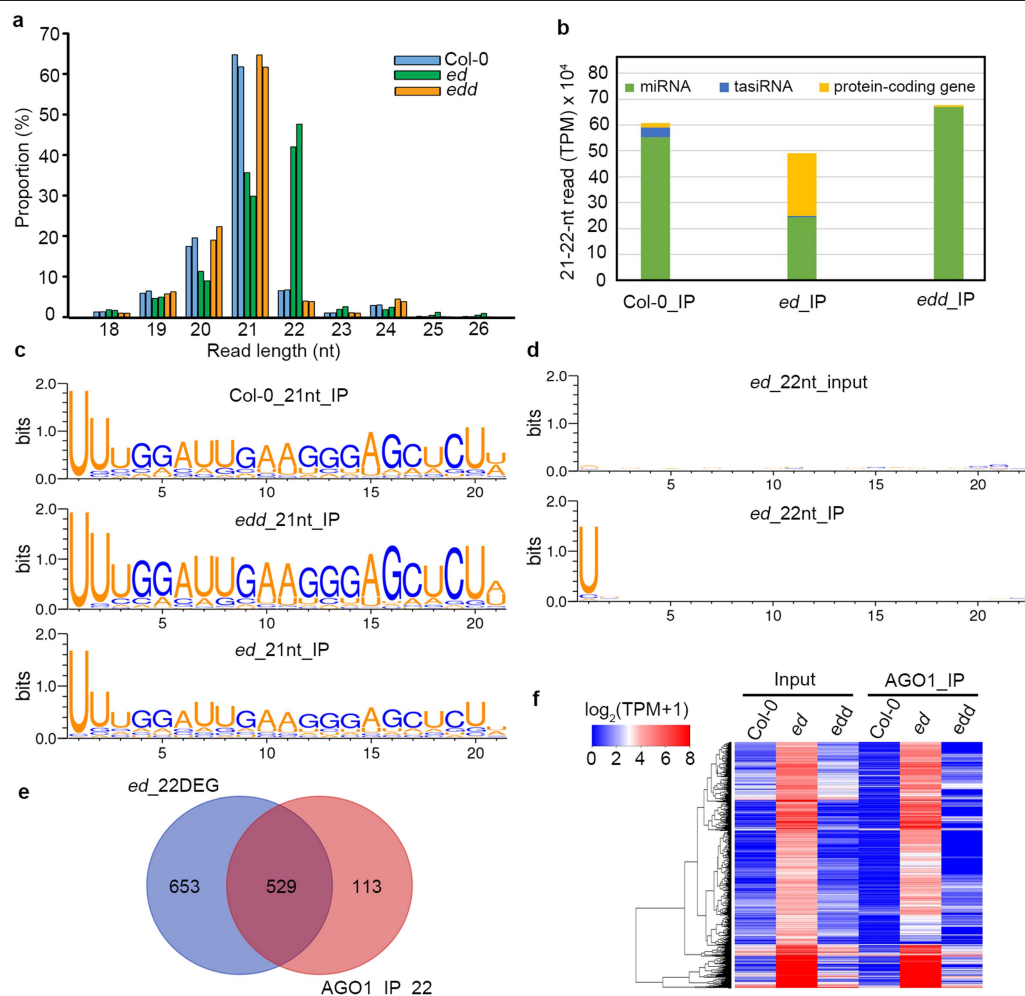
**Extended Data Fig. 3 | Accumulation of siRNAs and polysome profiles of the *ein5-1 ski2-3* mutant.** **a**, Representative images of 20-day-old Col-0, *ein5-1 ski2-3* and *ein5-1 ski2-3 dcl4-2 dcl2-1* plants. Scale bar, 2 cm. Experiments were repeated three times with similar results. **b**, Normalized abundance of 21- and 22-nt siRNAs from all genes, excluding miRNAs and tasiRNAs, in each genotype. Data shown as means  $\pm$  s.d.;  $n = 4$  biological independent samples. **c**, IGV of 21- and 22-nt siRNAs from representative *NIA1/2* gene loci in 20-day-old plants of each genotype. **d**, Polysome profiling of global translation efficiency in 20-day-old *ein5-1 ski2-3* plants; 10–50% sucrose gradient absorbance

(at 260 nm) was monitored in different fractions. Experiments were repeated twice with similar results. **e–g**, Polysome distribution of *NIA1/2* transcripts, with numbers on x-axis corresponding to fractions shown in **d**. *NIA1/2* mRNA levels were normalized against *PP2AA3* (AT1G13320). In **e**, **f**, *NIA1/2* expression in each polysomal fraction was calculated as the percentage of its expression in total RNA. In **g**, *TUB2* was used as a control. Data in **e–g** are presented as means and individual data points (technical replicates). Experiments were repeated twice with similar results.



**Extended Data Fig. 4 | Genetic combinations of *ein5-1 dcl4-2*, *ski2-2 dcl4-2* and sRNA-pathway-component mutants.** **a**, Western blot showing NIA1/2 protein levels at 15:00 in 20-day-old plants of the indicated genotypes. HSP90 was used as a control. **b–g**, Genetic interaction of *ein5-1 dcl4-2* and *ski2-2 dcl4-2* with *ago2-1* (**b**, **c**), *ago4-1* (**b**, **c**), *hen1-8* (**d**, **e**) and *amp1-30* (**f**, **g**). Representative images of 20-day-old plants of the indicated genotypes. Scale bars, 2 cm. **h**, NR

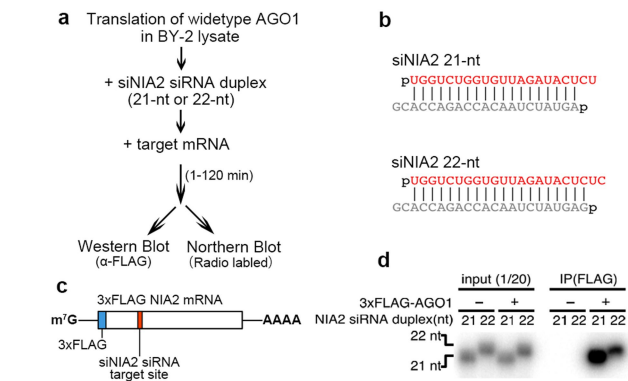
protein levels (anti-NR antibody recognizes both NIA1 and NIA2) in 20-day-old plants of the indicated genotypes. HSP90 was used as a control. **a**, **h**, Numbers to the right of gels indicate the molecular mass (in kDa) of proteins. The numbers of individual biological experiments are as follows: **a**, **h**,  $n = 2$ ; **b–g**,  $n = 3$ . For gel source data, see Supplementary Fig. 1.



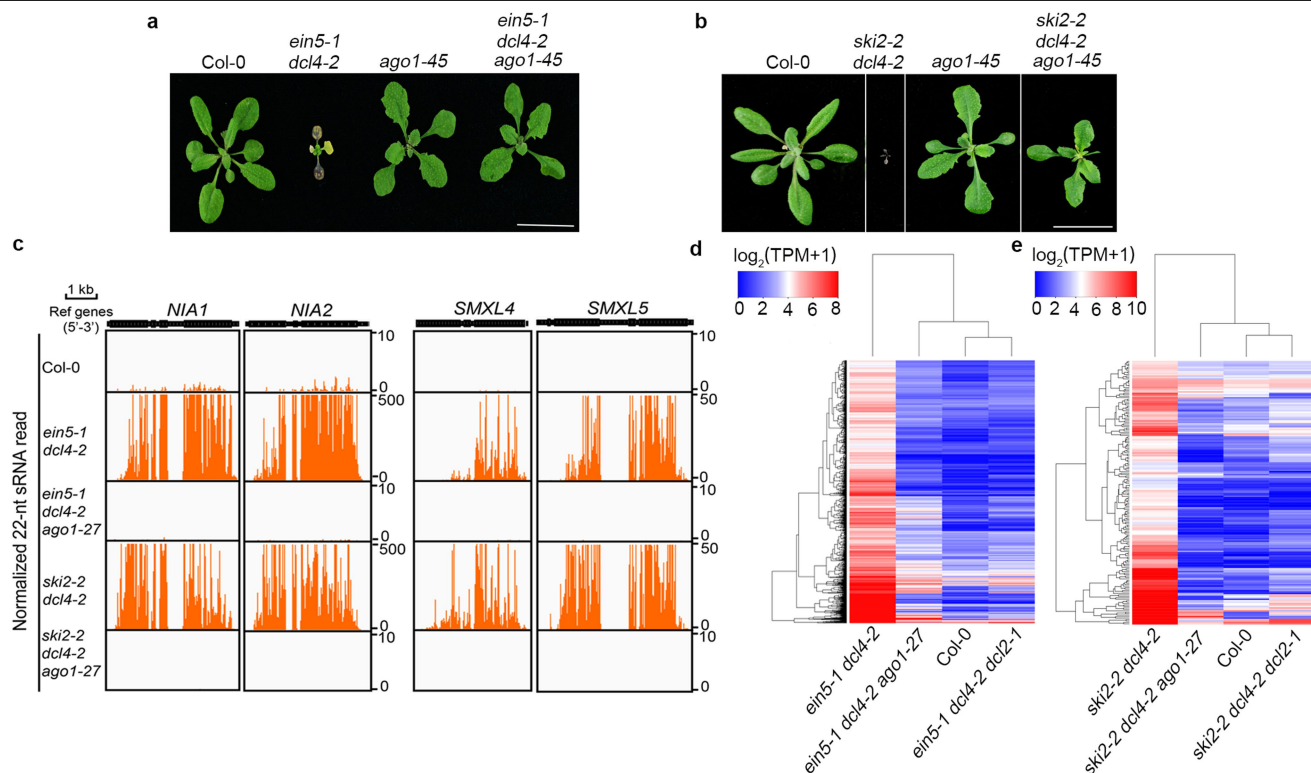
**Extended Data Fig. 5 | AGO1 associates with 22-nt sRNAs in *ein5-1 dcl4-2* plants. a**, Length distribution (18–26 nt) of AGO1-associated sRNAs in Col-0, *ein5-1 dcl4-2* (*ed*) and *ein5-1 dcl4-2 dcl2-1* (*edd*) plants. **b**, Classification (miRNAs, tasiRNAs and protein-coding-gene-derived siRNAs) and abundance of AGO1-associated sRNAs in Col-0 (Col-0\_IP), *ein5-1 dcl4-2* (*ed\_IP*) and *ein5-1 dcl4-2 dcl2-1* (*edd\_IP*). The abundance of 21- and 22-nt sRNAs only was calculated. **c**, **d**, Sequence conservation analysis of AGO1-associated 21-nt (**c**) or

22-nt (**d**) sRNA using Weblogo 3 software. **e**, Venn diagram shows the overlap between 1,182 22-nt-siRNA-generating genes in *ein5-1 dcl4-2* (*ed\_22DEG*) and 642 AGO1-associated 22-nt-siRNA-generating genes in *ein5-1 dcl4-2* (AGO1\_IP\_22). **f**, Heat map depicting the sRNA abundance ( $\log_2(\text{TPM} + 1)$ ) of 1,182 22-nt-siRNA-generating genes from *ein5-1 dcl4-2* plants in total (Input) and AGO1-immunoprecipitated (AGO1\_IP) sRNA sequencing experiments.



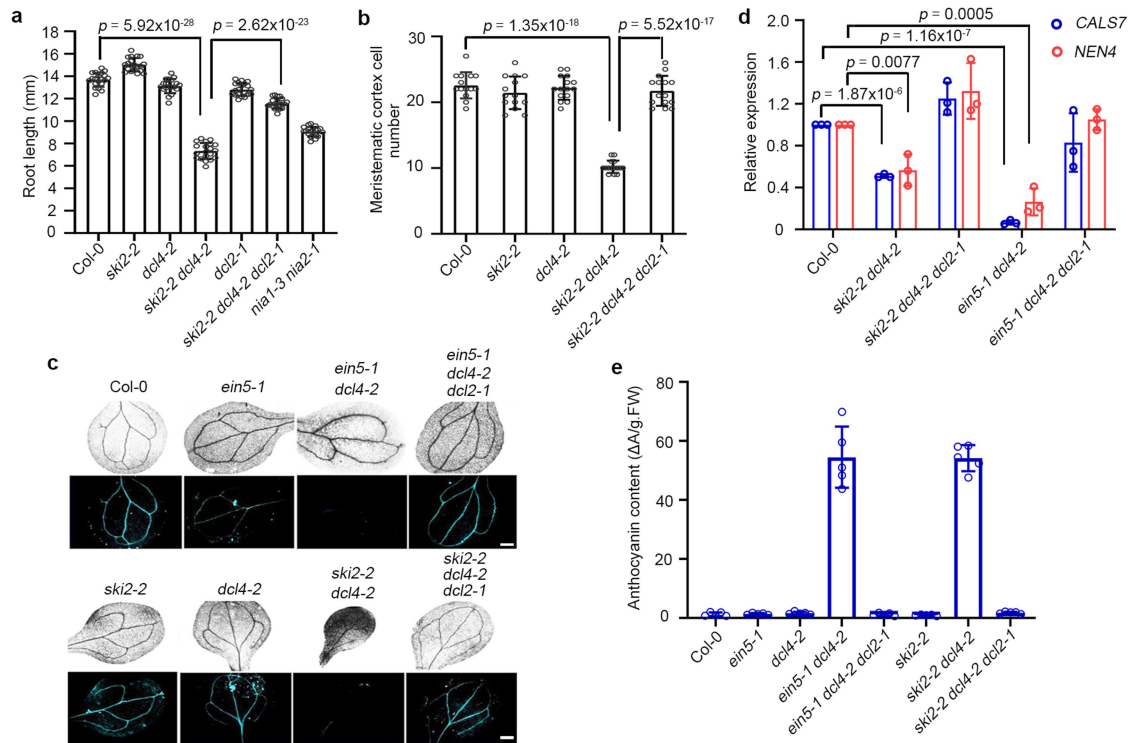


**Extended Data Fig. 6 | Scheme for RNA silencing by 21- and 22-nt siRNAs in vitro.** **a**, Scheme underlying the experimental procedure of Fig. 3d. See Methods for further details. **b**, Structures of 21- and 22-nt siNIA2 siRNAs. The guide strands shown in red were radiolabelled. The grey strands show the passenger strand. The 5'-ends of siRNAs were phosphorylated and the 3'-ends were methylated. **c**, Schematic of 3× *FLAG NIA2* mRNA and siNIA2 siRNA target site. **d**, We incubated 1.5 μM radiolabelled NIA2-siRNA duplexes (21- and 22-nt) with in vitro translated 3× FLAG-AGO1 in BY-2 lysate at 25 °C for 90 min. AGO1-RISC was then isolated using anti-FLAG antibody. Input (1/20) and AGO1-associated (IP (FLAG)) 21- and 22-nt siRNAs were then separated on denaturing gels and analysed by autoradiograph. Experiments were repeated twice with similar results. For gel source data, see Supplementary Fig. 1.



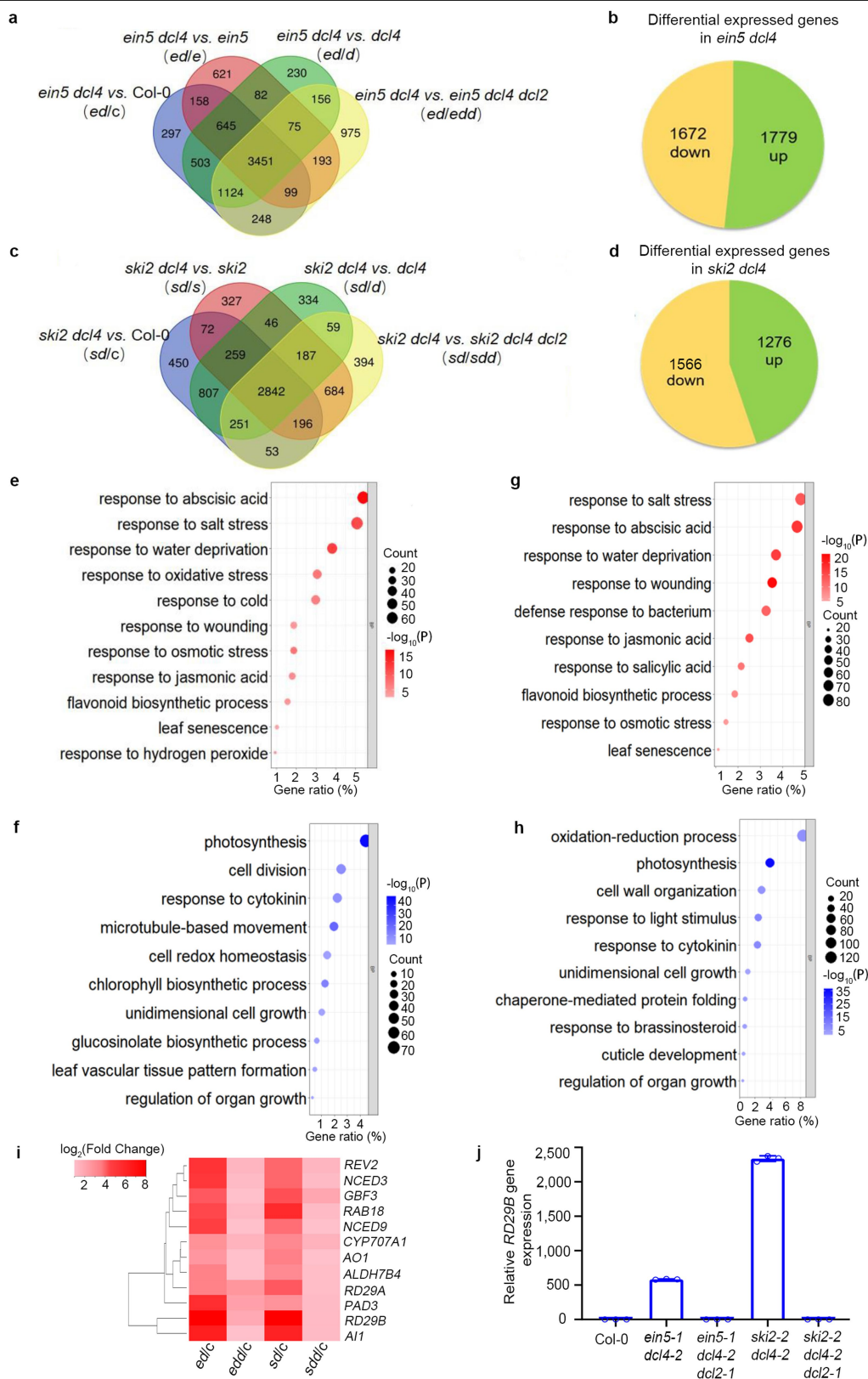
**Extended Data Fig. 7 | The *ago1* mutant rescues the growth defects and eliminates the accumulation of 22-nt siRNAs in *ein5-1 dcl4-2* and *ski2-2 dcl4-2* plants.** **a, b**, Representative images of 20-day-old plants. Scale bars, 2 cm. Experiments were repeated three times with similar results. **c**, IGV of 22-nt

siRNAs derived from *NIA1/2* and *SMXL4/5* gene loci in each genotype of 20-day-old plants. **d, e**, Heat map showing the abundance of 22-nt siRNAs ( $\log_2(\text{TPM} + 1)$ ) from 1,182 (from *ein5-1 dcl4-2* plants) and 182 (from *ski2-2 dcl4-2* plants) genes that produce 22-nt siRNAs in the indicated genotypes.



**Extended Data Fig. 8 | Root growth and phloem formation are impaired in *ein5-1 dcl4-2* and *ski2-2 dcl4-2* plants. **a**, Root length of 8-day-old plants of indicated genotypes. Data are means  $\pm$  s.d.;  $n = 20$  biologically independent roots. **b**, Number of meristematic cortex cells in 9-day-old plants. Data are means  $\pm$  s.d.;  $n = 14$  (for Col-0 and *ski2-2*) and 15 (for *dcl4-2*, *ski2-2 dcl4-2* and *ski2-2 dcl4-2 dcl2-1*) biologically independent roots. **c**, Cotyledons from 18-day-old plants shown under bright field or with aniline-based callose visualization. Scale bars, 500  $\mu$ m.  $n = 5$  biologically independent cotyledons with similar**

results. See Methods for further details. **d**, Relative expression of phloem-related genes (*CALS7* and *NEN4*) in the indicated genotypes of 20-day-old plants. Data are means  $\pm$  s.d.;  $n = 3$  biologically independent experiments. **e**, Measurement of relative anthocyanin content ( $\Delta A/g$  FW: absorbance at 525 nm minus absorbance at 650 nm per gram of fresh weight) in 20-day-old plants of the indicated genotypes. Data are means  $\pm$  s.d.;  $n = 5$  biologically independent samples.  $P$ -values in **a**, **b**, **d** are from two-tailed Student's  $t$ -tests.

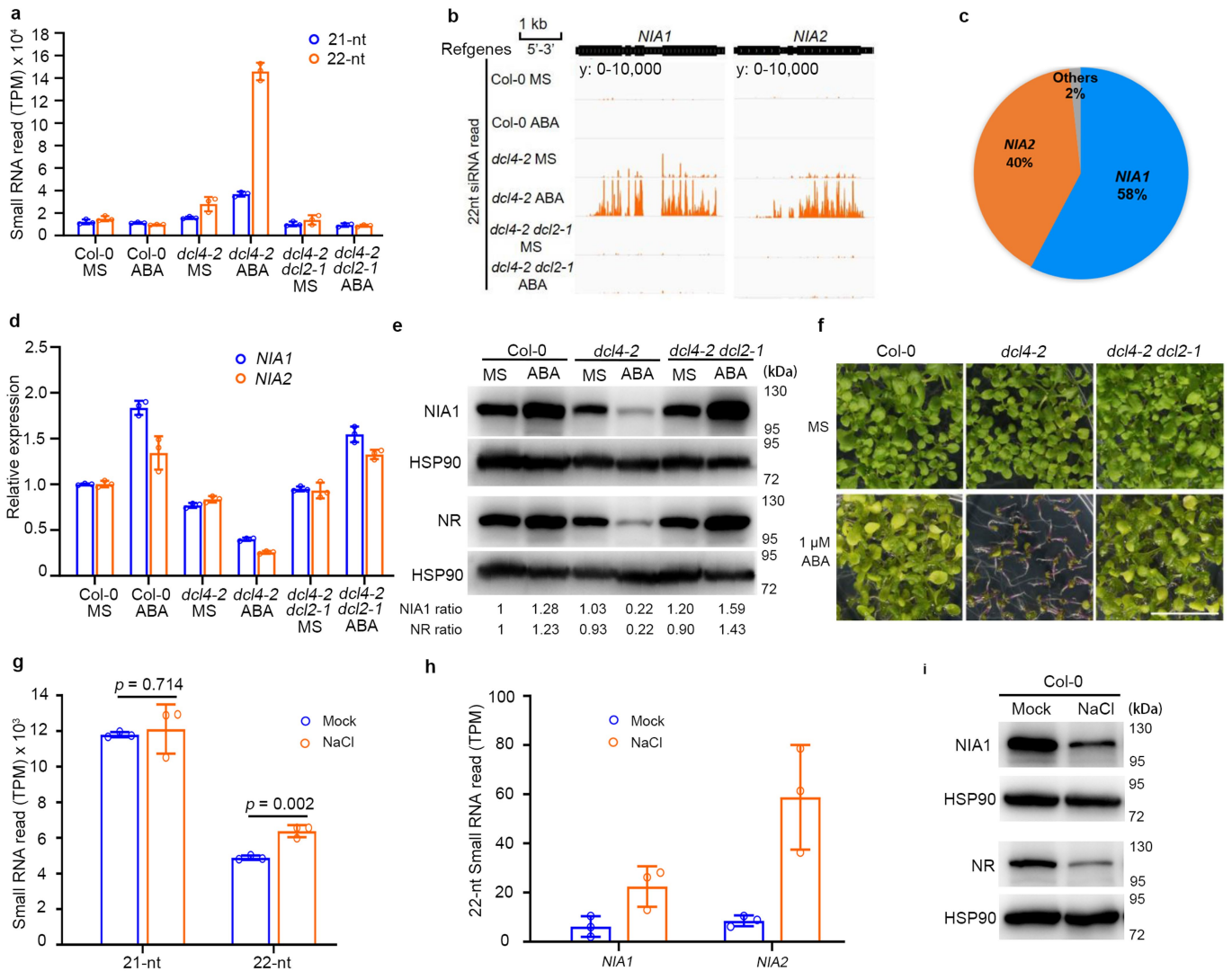


**Extended Data Fig. 9** | See next page for caption.



**Extended Data Fig. 9 | Transcriptome profiles and gene-ontology analysis of differentially expressed genes in *ein5-1 dcl4-2* and *ski2-2 dcl4-2* plants.**

**a, c,** Venn diagrams depicting numbers of DCL2-dependent differentially expressed genes (at least twofold changes, FDR < 0.01) in *ein5-1 dcl4-2* (**a**) and *ski2-2 dcl4-2* (**c**) plants. **b, d,** Pie charts representing the numbers of upregulated and downregulated genes in *ein5-1 dcl4-2* (**b**) and *ski2-2 dcl4-2* (**d**) plants. **e–h,** Representative gene-ontology enrichment analysis of upregulated (red) and downregulated (blue) genes in *ein5-1 dcl4-2* (**e, f**) and *ski2-2 dcl4-2* (**g, h**) plants. Circle sizes represent gene numbers and colour gradients indicate enrichment significance. Numbers at the bottom are the ratios of genes for each indicated biological process in the total differentially expressed genes. **i,** Heatmap depicting the log<sub>2</sub>(fold change) of representative genes involved in ABA biogenesis and response in 20-day-old Col-0 (**c**) versus *ein5-1 dcl4-2* (*ed*), *ein5-1 dcl4-2 dcl2-1* (*edd*), *ski2-2 dcl4-2* (*sd*) and *ski2-2 dcl4-2* (*sdd*) plants. **j,** qRT-PCR analysis of an ABA-responsive gene (*RD29B*) in 20-day-old plants of the indicated genotypes; data shown are means ± s.d. FDR values used to filter the differentially expressed genes in **a–h** were obtained by Benjamini–Hochberg analysis; *n* = 3 biologically independent samples.



**Extended Data Fig. 10 | ABA and salt treatments induce the production of 22-nt siRNAs from *NIA1* and *NIA2* loci.** **a**, Normalized abundance of 21- and 22-nt siRNAs from all genes, excluding miRNAs and tasiRNAs, in each condition. **b**, Accumulation of 22-nt siRNAs from *NIA1/2* gene loci in 14-day-old plants under 1  $\mu$ M ABA treatment. **c**, Pie chart showing percentages of loci generating 22-nt siRNAs under 1  $\mu$ M ABA treatment. **d**, Relative *NIA1/2* mRNA expression level at 15:00 in 14-day-old plants on MS medium or 1  $\mu$ M ABA medium. **e**, *NIA1/2* protein levels at 15:00 in 14-day-old plants on MS medium or 1  $\mu$ M ABA medium. HSP90 was used as a control to normalize the protein level. **f**, Representative images of 14-day-old seedlings grown on the indicated medium. Scale bar, 5 mm. **g**, Normalized abundance of 21- and 22-nt siRNAs

from all genes, excluding miRNAs and tasiRNAs, in 15-day-old Col-0 plants under 250 mM NaCl treatment (data retrieved from GEO, accession number GSE66599). All *P*-values are from two-tailed Student's *t*-tests. **h**, Normalized abundance of 22-nt siRNAs from *NIA1* and *NIA2* genes under the same conditions as in **g**. **i**, *NIA1/2* protein level detected by western blot in 8-day-old wild-type seedlings upon 250 mM NaCl treatment for 24 h. HSP90 was used as a control. **e**, **i**, Numbers to the right of gels indicate the molecular mass (in kDa) of proteins. Data are presented as means  $\pm$  s.d. in **a**, **d**, **g**, **h**. The numbers of individual biological experiments are as follows: **a**, **d**, **f**,  $n = 3$ ; **e**,  $n = 2$ ; **i**,  $n = 4$ . For gel source data, see Supplementary Fig. 1.

## Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

### Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- ☐ ☒ The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement
- ☐ ☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- ☐ ☒ The statistical test(s) used AND whether they are one- or two-sided  
*Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
- ☒ ☐ A description of all covariates tested
- ☐ ☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- ☐ ☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- ☐ ☒ For null hypothesis testing, the test statistic (e.g.  $F$ ,  $t$ ,  $r$ ) with confidence intervals, effect sizes, degrees of freedom and  $P$  value noted  
*Give  $P$  values as exact values whenever suitable.*
- ☒ ☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- ☒ ☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- ☒ ☐ Estimates of effect sizes (e.g. Cohen's  $d$ , Pearson's  $r$ ), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

### Software and code

Policy information about [availability of computer code](#)

#### Data collection

Triax v1.44A(polysome profiling);  
Tanon 5200 (Western Blot assay);  
Typhoon FLA 7000 (PhosphorImaging);  
ImageJ 1.46r(Root length measurement and Western blot/Northern blot quantification);  
LightCycler 96 v1.1.0.1320 (qRT-PCR assay);  
Zen Black (Confocal assay);  
Zen 2.0 Pro (Phloem observation).

#### Data analysis

We fully described the software availability in the manuscript as:  
For mRNA-seq: Raw reads were filtered by Trimmomatic (v0.38) with default parameter. Then filtered reads were aligned to the Arabidopsis Col-0 genome (TAIR10 release) using tophat2 (v2.1.1) program with 2 mismatches allowed. The differential expressed genes were called by cuffdiff (v2.2.1).  
  
For sRNA-seq: Raw reads were filtered by fastx\_toolkit (v0.0.14). The 18-26 nt clean reads were mapped to structural RNA database. Then the unmapped reads were aligned to TAIR10 genome using bowtie (v1.2.2) program with following parameters: -v 0 -a -M 3 --best --strata. Differential small RNA-generating genes were called by DESeq package in R. GO enrichment analysis was performed in DAVID Bioinformatics Resources 6.8 and the BiNGO package in Cytoscape v3.4.0 software with default parameters. Mapped reads were visualized in the Integrated Genome Viewer, version IGV\_2.3.72; ShortStack software (v3.8.5) were used for small RNA phase analysis.  
  
Weblogo 3.6.0 was used for small RNA sequence conservation analysis;  
  
Graphpad Prism 8.0.2 was used for graphs data presentation;

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

## Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Sequencing data are available to public at the NCBI Gene Expression Omnibus (GEO) repository under the accession number GSE136164. Source gel data for immunoblots and radiograms (Fig. 2-4, Extended data Fig. 2,4, 6,10) are supplied as Supplementary Figure 1; Source data for all Graphs (Fig.1-4, Extended data Fig.1-3,5,7-10) are also provided and are available with the online version of the paper. All other datasets generated within this study available from the corresponding author upon reasonable request.

## Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://nature.com/documents/nr-reporting-summary-flat.pdf)

## Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No statistical methods were used to predetermine sample sizes. Required experimental sample sizes were estimated based on previous established protocols in the field. The sample sizes were adequate as the experimental results were reproducible.
Data exclusions	No data were excluded from the analysis.
Replication	All experimental findings were reproduced in several independent biological experiments (N) with multiple technical replicates. The number of repeats is indicated in the figure legends. Main Conclusions were confirmed in different assays (Immunoblot assay, genetic assay, In vitro cleavage assay) handled by multiple researchers and across different labs.
Randomization	Samples of the same genotypes were randomly collected and pooled for downstream experiments.
Blinding	Blinding was not possible as the author who performed the experiments also analyzed the data.

## Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

### Materials & experimental systems

n/a	Involved in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data

### Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

## Antibodies

Antibodies used	anti-NIA1(Provided by ABClonal Biotechnology, Cat. WG-03170, dilution 1:2,000 ); anti-NR(Provided by Agrisera, Cat. AS08 310, dilution 1:2,000); anti-AGO1(Provided by Agrisera, Cat. AS09 527, dilution 1:300 for AGO1-IP); anti-HSP90(Provided by Beijing Protein Innovation, Cat. AbM51099-31-PU, dilution 1:400,000); anti-Tubulin(Provided by Abclonal, Cat. AC010, dilution 1:5,000); anti-Histone H3(Provided by Abclonal, Cat. A2348,dilution 1:5,000).
Validation	1. anti-NIA1 antibody made by this study was validated in Arabidopsis for Western blot and the validated data was presented in the manuscript Extended Data Fig.2e; Concerning other commercial antibody specificity, we kindly refer to the supplier's websites and datasheets to find statements on specificity and citations for the use of the antibodies:



2. The commercial anti-NR antibody has been well validated in Arabidopsis for Western Blot in the several published papers and in our manuscript (Extended data fig.2e), and the antibody profile can be found in the following link: <https://www.agrisera.com/en/artiklar/nr-nitrate-reductase-assimilatory.html>;
3. The commercial anti-AGO1 antibody has been well validated in Arabidopsis for Western blot and Small RNA-IP sequencing in the published papers, and the antibody profile can be found in the following link: <https://www.agrisera.com/en/artiklar/ago1-argonaute-1.html>;
4. The commercial anti-HSP90 antibody has been well validated in Arabidopsis for Western blot, and the antibody profile can be found in the following link: <http://www.proteomics.org.cn/nav/108.html>.
5. The commercial anti-Tubulin antibody has been well validated in Arabidopsis for Western blot in the published papers,, and the antibody profile can be found in the following link: <https://abclonal.com/instructions/pdf/AC010>;
6. The commercial anti-Histone H3 antibody has been well validated in Arabidopsis for Western blot in the published papers,, and the antibody profile can be found in the following link: <https://abclonal.com/instructions/pdf/A2348>;

# Childhood vaccines and antibiotic use in low- and middle-income countries

<https://doi.org/10.1038/s41586-020-2238-4>

Received: 3 December 2019

Accepted: 5 March 2020

Published online: 29 April 2020

Open access

 Check for updates

Joseph A. Lewnard<sup>1,2,3✉</sup>, Nathan C. Lo<sup>4</sup>, Nimalan Arinaminpathy<sup>5</sup>, Isabel Frost<sup>5,6</sup> & Ramanan Laxminarayan<sup>6,7</sup>

Vaccines may reduce the burden of antimicrobial resistance, in part by preventing infections for which treatment often includes the use of antibiotics<sup>1–4</sup>. However, the effects of vaccination on antibiotic consumption remain poorly understood—especially in low- and middle-income countries (LMICs), where the burden of antimicrobial resistance is greatest<sup>5</sup>. Here we show that vaccines that have recently been implemented in the World Health Organization’s Expanded Programme on Immunization reduce antibiotic consumption substantially among children under five years of age in LMICs. By analysing data from large-scale studies of households, we estimate that pneumococcal conjugate vaccines and live attenuated rotavirus vaccines confer 19.7% (95% confidence interval, 3.4–43.4%) and 11.4% (4.0–18.6%) protection against antibiotic-treated episodes of acute respiratory infection and diarrhoea, respectively, in age groups that experience the greatest disease burden attributable to the vaccine-targeted pathogens<sup>6,7</sup>. Under current coverage levels, pneumococcal and rotavirus vaccines prevent 23.8 million and 13.6 million episodes of antibiotic-treated illness, respectively, among children under five years of age in LMICs each year. Direct protection resulting from the achievement of universal coverage targets for these vaccines could prevent an additional 40.0 million episodes of antibiotic-treated illness. This evidence supports the prioritization of vaccines within the global strategy to combat antimicrobial resistance<sup>8</sup>.

Antimicrobial resistance (AMR) poses a substantial and growing threat to global health and economic well-being<sup>9</sup>. Populations of LMICs suffer the greatest morbidity and mortality associated with AMR, and stand to be disproportionately affected by projected increases in this burden over future decades<sup>10</sup>. Antibiotic use is a key risk factor for human colonization or infection with drug-resistant organisms at scales spanning individuals, communities and countries<sup>11</sup>. Increasing rates of antibiotic consumption in LMICs<sup>12</sup> have highlighted tension between the need to ensure access to life-saving antibiotics globally, and the need to curb resistance selection<sup>13</sup>. Although it has been widely hypothesized that vaccines may reduce AMR burden by preventing infections that drive antibiotic consumption<sup>1</sup>, few real-world studies provide evidence of this effect<sup>14</sup>.

Acute respiratory infection (ARI) and diarrhoea are the leading causes of antibiotic use among children in LMICs<sup>15</sup>. Because few diagnostic tools exist to guide management of ARI and diarrhoea, antibiotic treatment of these conditions is generally informed by suspicion rather than confirmation of disease aetiology. Pneumococcal conjugate vaccines (PCVs) against 10 and 13 serotypes of *Streptococcus pneumoniae* and live attenuated rotavirus vaccines target the predominant causes of ARI and diarrhoea among children, respectively<sup>6,16</sup>. These vaccines have recently been introduced into routine paediatric immunization programmes of countries across the globe<sup>17</sup>. Although appropriate

for children experiencing pneumococcal ARI, antibiotic treatment is not warranted for rotavirus-attributable diarrhoea. To date, no studies have addressed the impact of PCV10/13 and rotavirus vaccines on antibiotic use in LMICs<sup>5</sup>.

We analysed data from demographic and health surveys to estimate: (i) the incidence of antibiotic treatment for ARI and diarrhoea among children in LMICs; (ii) the effectiveness of PCV10/13 and rotavirus vaccines in preventing antibiotic use associated with ARI and diarrhoea among children in these settings; and (iii) the proportion of antibiotic-treated ARI and diarrhoea cases attributable to vaccine-serotype *S. pneumoniae* and rotavirus, respectively. We used these results to quantify the incidence of antibiotic consumption among children in LMICs prevented by current uses of PCV10/13 and rotavirus vaccines, and the additional consumption that could be averted by achieving universal vaccine coverage targets<sup>18</sup>.

## Vaccine effects against antibiotic use

We first assessed whether PCV10/13 and rotavirus vaccines are effective in preventing antibiotic-treated illnesses among children in LMICs. We undertook a case–control study comparing the odds of previous vaccination among children under five years of age reported to have experienced (cases) or not to have experienced (controls) ARI and

<sup>1</sup>Division of Epidemiology, School of Public Health, University of California, Berkeley, Berkeley, CA, USA. <sup>2</sup>Division of Infectious Diseases and Vaccinology, School of Public Health, University of California, Berkeley, Berkeley, CA, USA. <sup>3</sup>Center for Computational Biology, College of Engineering, University of California, Berkeley, Berkeley, CA, USA. <sup>4</sup>Department of Medicine, University of California, San Francisco, San Francisco, CA, USA. <sup>5</sup>School of Public Health, Imperial College London, London, UK. <sup>6</sup>Center for Disease Dynamics, Economics & Policy, New Delhi, India.

<sup>7</sup>Princeton Environmental Institute, Princeton University, Princeton, NJ, USA. ✉e-mail: [jlwnard@berkeley.edu](mailto:jlwnard@berkeley.edu)

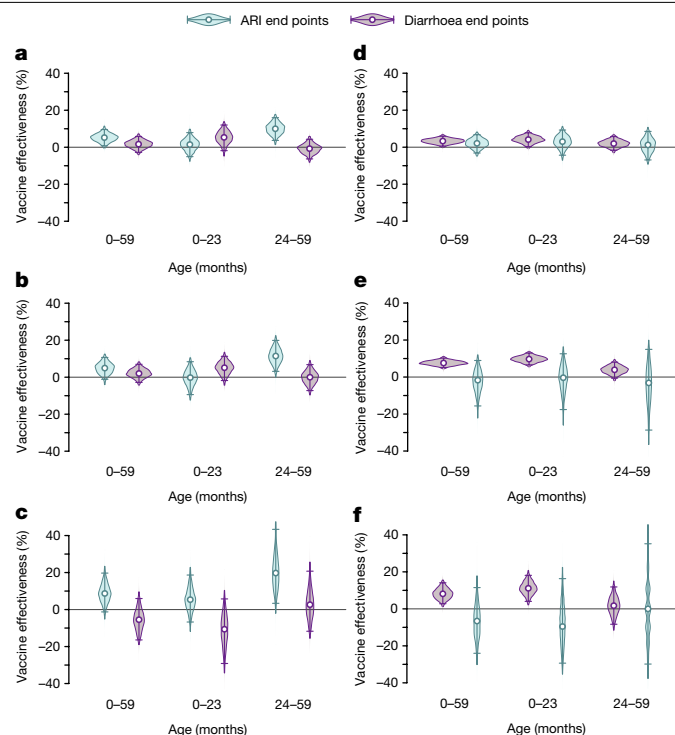
diarrhoea in the two weeks preceding the standardized household surveys. End points for our analysis were (i) all cases of disease, (ii) cases for which medical care or advice was sought, and (iii) cases receiving antibiotic treatment. Data were available on 65,815 children from surveys conducted from 2015 onward in Afghanistan, Angola, Armenia, Burundi, Ethiopia, Haiti, Lao People's Democratic Republic, Malawi, Nepal, Pakistan, the Philippines, Senegal, Sierra Leone, South Africa, Tajikistan, the United Republic of Tanzania, Uganda and Zimbabwe (Supplementary Tables 1, 2; earlier surveys did not collect children's history of PCV10/13 and rotavirus vaccination). To eliminate confounding effects driven by any association between the access of children to vaccines and the likelihood of disease or antibiotic treatment, we matched cases and asymptomatic controls from the same countries for age and survey date (each within one month), urbanicity, within-country wealth quintile and receipt of pentavalent vaccine doses (a potential confounder of ARI analyses, and a marker of vaccine and/or healthcare access). We also conducted planned subgroup analyses for children aged 0–23 months and aged 24–59 months due to well-documented differences in ARI and diarrhoea aetiology across age groups<sup>6,7</sup>.

Children who received at least three PCV10/13 doses experienced 8.7% lower odds of antibiotic-treated ARI than unvaccinated children (Fig. 1 and Supplementary Table 3), although the 95% confidence interval for this estimate included the possibility of no effect (–1.3–19.7% reduction (values in parentheses indicate 95% confidence intervals throughout); two-sided  $P=0.079$ ). The estimated reduction was driven primarily by the prevention of cases in children aged 24–59 months; within this age group, vaccination was associated with a 19.7% (3.4–43.4%) reduction in cases of antibiotic-treated ARI. By contrast, effects of PCV10/13 were not clearly evident during the first 2 years of life, when less than 5% of cases of ARI are attributable to vaccine-serotype pneumococci<sup>7</sup>. To test that our findings are the result of vaccine-conferred protection and not confounding factors, we also estimated PCV10/13 effectiveness against diarrhoea and antibiotic-treated diarrhoea. Although these outcomes would not be expected to be causally related to PCV10/13 receipt, their association with previous vaccination may be prone to similar confounding pathways. We did not find strong evidence of associations between negative-control outcomes and receipt of PCV10/13, validating our finding that PCV10/13-conferred protection reduces the incidence of antibiotic-treated ARI.

Similarly, children who received a full rotavirus vaccine series (at least two doses for countries included in this study) experienced 8.1% (2.8–14.1%) lower odds of antibiotic-treated diarrhoea at ages of 0–59 months than their unvaccinated counterparts (Fig. 1 and Supplementary Tables 4, 5). During the first 2 years of life, we estimated 11.4% (4.0–18.6%) effectiveness of rotavirus vaccination against antibiotic-treated diarrhoea. Protection was not evident in children aged 24–59 months, when rotavirus is a less common cause of diarrhoea among children in LMICs<sup>6</sup>. The absence of associations between rotavirus vaccination and negative-control ARI end points provided validation for our finding that vaccine-conferred protection against rotavirus reduces the risk to children of antibiotic-treated diarrhoea.

## Analysing aetiological fractions using vaccine effects

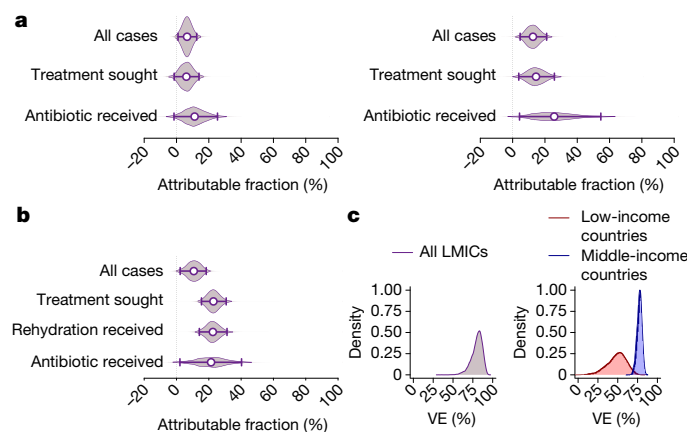
We next developed a simple model using the vaccine effectiveness estimates described above to investigate the proportions of antibiotic-treated ARI and diarrhoea attributable to vaccine-serotype pneumococci and rotavirus, respectively (Methods). The model took these inputs together with pooled estimates of vaccine efficacy against disease caused by the targeted pathogens based on meta-analyses of previous studies (Supplementary Table 7). Because rotavirus vaccine efficacy beyond the second year of life is uncertain (and likely to be low)<sup>19,20</sup>, and because we did not identify strong evidence of vaccine-conferred protection against antibiotic-treated diarrhoea in children aged 24–59 months, we limited our assessment of rotavirus-attributable antibiotic use to children aged 0–23 months.



**Fig. 1 | Effectiveness of pneumococcal and rotavirus vaccines against illness and antibiotic treatment.** **a–f**, The estimated effectiveness against ARI and diarrhoea end points of PCV10/13 (**a–c**) and rotavirus vaccines (**d–f**) for all cases (**a, d**), cases for which treatment or advice was sought (**b, e**) and cases that were treated with antibiotics (**c, f**). Estimates were calculated as one minus the matched odds ratio and are shown as vaccine effectiveness. Analyses matched children with each end point to asymptomatic controls on the basis of country, age (within 1 month), visit timing (within 1 month), wealth quintile (country-specific), urbanicity and pentavalent vaccine doses received. The population available for analysis included 5,342 ARI cases (of whom 3,294 sought treatment or advice and 1,913 received antibiotics) and 57,856 controls without ARI; and 9,944 diarrhoea cases (of whom 7,382 sought treatment or advice and 1,437 received antibiotics) and 40,059 controls without diarrhoea (Supplementary Tables 1, 2). Points and lines indicate median estimates and 95% confidence intervals, respectively. We estimated vaccine effectiveness against negative-control end points (PCV10/13 effect against diarrhoea; rotavirus vaccine effect against ARI) to assess residual confounding as a validation step. PCV10/13 exposure was defined as  $\geq 3$  doses received. Because all countries in this analysis used Rotarix in their national immunization program, we defined  $\geq 2$  doses as a full rotavirus vaccination series. Numerical estimates can be found in Supplementary Tables 3–5. Quantiles are estimated through 2,000 independent draws from the distribution of estimates.

We estimated that vaccine-serotype pneumococci would account for 10.9% (–1.6–25.4%) of antibiotic-treated ARI cases among children under five years old in PCV10/13-naïve LMIC populations (Fig. 2 and Supplementary Table 8); the negative lower limit of this confidence interval resulted from propagation of uncertainty in the estimated effect of PCV10/13 against antibiotic-treated ARI in children aged 0–59 months, and reflects the possibility that vaccine-targeted pneumococcal serotypes are not a significant cause of antibiotic consumption. Among children aged 24–59 months, we estimated that vaccine-serotype pneumococci cause 24.8% (4.3–54.6%) of antibiotic-treated ARI cases.

We similarly estimated that rotavirus would cause 21.6% (2.1–40.3%) of antibiotic-treated diarrhoea among children aged 0–23 months in unvaccinated LMIC populations. Consistent with the findings of diarrhoea aetiology studies undertaken across differing LMICs<sup>6,21</sup>, we obtained similar estimates of the rotavirus-attributable fraction in analyses stratified by the income status of each country, in which we



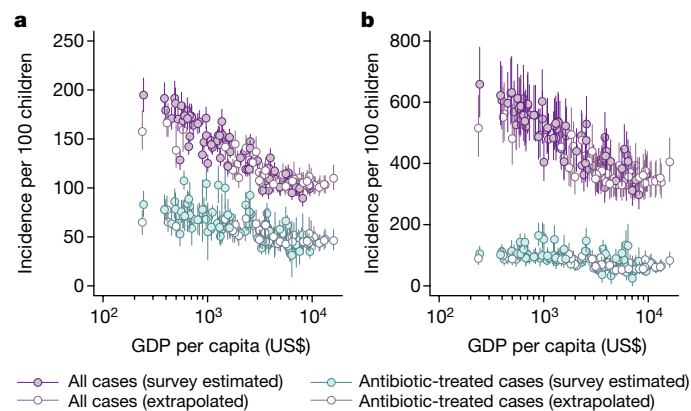
**Fig. 2 | Estimates of the attributable fraction for vaccine-preventable infections.** **a, b,** We illustrate estimates of pathogen-specific attributable fractions of vaccine-serotype pneumococci for children aged 0–59 months (**a**, left) and 24–59 months (**a**, right), and rotavirus for children aged 0–23 months (**b**). **c,** We also illustrate distributions of vaccine efficacy estimates against infections involving the vaccine-targeted organism, as estimated using meta-analyses (Supplementary Table 7); these estimates provided a basis for computing the attributable fraction (Supplementary Tables 1, 2). We considered PCV efficacy against vaccine-serotype invasive pneumococcal disease for the primary analysis (left). A secondary analysis (Supplementary Table 8) used PCV efficacy against culture-confirmed pneumococcal vaccine-serotype acute otitis media. For rotavirus (right), we stratified estimates of human monovalent rotavirus vaccine (Rotarix) efficacy against rotavirus gastroenteritis in children aged 0–23 months in middle-income countries and low-income countries to account for the differential efficacy in these settings, and obtained a pooled estimate of the attributable fraction weighted by the number of children residing in middle-income countries and low-income countries (Methods). VE, vaccine effectiveness. Points and lines indicate median estimates and 95% confidence intervals, respectively. Quantiles are obtained through 2,000 independent draws from the distribution of estimates.

accounted for variation in rotavirus vaccine efficacy across settings (Methods and Supplementary Tables 7, 8). We estimated that 21.3% (0.0–42.4%) and 22.7% (4.5–60.6%) of antibiotic-treated diarrhoea cases would be attributable to rotavirus among unvaccinated children in middle-income and low-income countries, respectively.

### Antibiotic use in LMICs

To understand these estimates of vaccine effectiveness and pathogen-attributable fractions in terms of absolute antibiotic consumption, we estimated the incidence of ARI-related antibiotic use among children aged 24–59 months, and of diarrhoea-related antibiotic use among children aged 0–23 months. We analysed data from household surveys covering 944,173 children across 77 countries from 2006 to 2018 (Methods). We used the resulting estimates to extrapolate country-level incidence rates for 58 additional countries without recent household surveys, aided by data on 405 national-level health, socioeconomic and demographic economic indicators from the World Bank Open Data catalogue (<https://data.worldbank.org/>) (Supplementary Tables 9, 10).

Estimated incidence rates of all cases and antibiotic-treated cases of ARI and diarrhoea were the highest in the poorest countries, and the lowest in the wealthiest countries (Fig. 3). Our estimates of country-specific ARI incidence in children aged 24–59 months ranged from 89.5 to 194.8 episodes per 100 children annually across all LMICs (Fig. 3 and Supplementary Table 11; we present further estimates of antibiotic-treated ARI in children aged 0–59 months in Supplementary Table 12). For children aged 24–59 months, the estimated proportion of



**Fig. 3 | Estimated incidence across countries of ARI and diarrheal illnesses per 100 children.** **a, b,** We estimate country-specific incidence of ARI and antibiotic-treated ARI in children aged 24–59 months (**a**) and diarrhoea and antibiotic-treated diarrhoea in children aged 0–23 months (**b**). Points and lines indicate median estimates for an individual country together with accompanying 95% confidence intervals. Estimates are obtained from analyses of Demographic Health Surveys and Multiple Indicator Cluster Surveys, comprising 944,173 children across 77 countries, as well as extrapolations based on 405 health, nutrition and population indicators for all LMICs. Points plotted in white are extrapolated from estimates based on household survey data (see Methods). Estimates for individual countries are provided in Supplementary Tables 11–13. Quantiles are obtained through 5,000 independent draws from the distribution of estimates.

episodes treated with antibiotics ranged from 30.0% to 69.4%, yielding incidence rates of antibiotic-treated ARI of between 31.2 and 107.3 cases per 100 children annually. For diarrhoea in children aged 0–23 months, we estimated that country-specific incidence rates ranged from 297.5 to 658.4 episodes per 100 children annually. Accounting for antibiotic treatment in an estimated 7.7–37.7% of diarrhoea episodes, annual rates of antibiotic-treated diarrhoea ranged from 24.6 to 164.5 episodes per 100 children. Country-specific estimates are included in Supplementary Tables 11–19.

### Pathogen-attributable antibiotic use

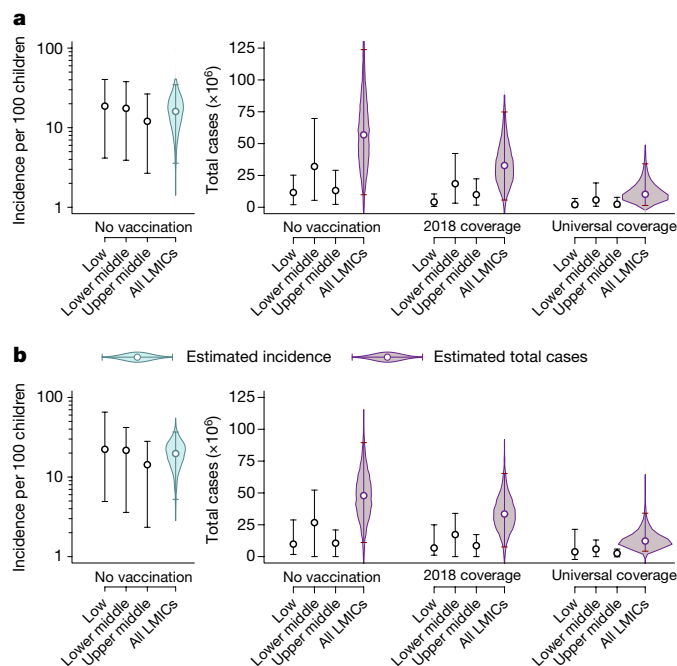
Multiplying these incidence rate estimates by pathogen-specific attributable fractions provides the burden of antibiotic-treated ARI and diarrhoea targeted by PCV10/13 and rotavirus vaccines. In the absence of vaccination, we estimated that vaccine-serotype pneumococci would cause 15.9 (2.8–34.7) antibiotic-treated ARI episodes per 100 children aged 24–59 months annually in LMICs, translating to a total of 56.9 (9.9–123.8) million antibiotic-treated ARI episodes (Fig. 4 and Supplementary Table 20). We estimated that 11.7 (2.0–24.8) million episodes occur annually in countries in which PCV10/13 has not been implemented. Analyses comprising all children aged 0–59 months residing in LMICs yielded similar burden estimates (Supplementary Table 21).

We estimated that rotavirus would cause 19.7 (4.6–36.8) antibiotic-treated diarrhoea episodes annually per 100 children aged 0–23 months in LMICs in the absence of vaccination, resulting in a total of 47.9 (11.1–89.5) million antibiotic-treated diarrhoea episodes (Fig. 4 and Supplementary Table 22). Of this burden, we estimated that 20.4 (3.4–38.7) million episodes occur annually in settings without rotavirus vaccination.

### Vaccine-preventable antibiotic use

The ‘direct’ protection conferred by the immunological effects of vaccines on their recipients provides a lower-bound estimate of the impact achievable by immunization programmes, as it excludes ‘indirect’





**Fig. 4 | Total vaccine-preventable antibiotic consumption and incidence per 100 children.** **a, b.** We estimated the incidence and total number of antibiotic-treated ARI and diarrhoea episodes attributable to PCV10/13-serotype pneumococci in children aged 24–59 months (**a**) and rotavirus in children aged 0–23 months (**b**), respectively. Left, incidence in the absence of vaccination. Right, the corresponding total number of cases, the total number of cases under 2018 vaccine coverage levels and under universal vaccine coverage. Estimates were stratified by income status (low income; lower middle income; upper middle income). Points indicate median estimates, with superimposed lines indicating 95% confidence intervals; violin plots illustrate the distribution around estimated incidence and total cases. Numerical estimates are provided in Supplementary Tables 20–22, 24–26. Quantiles are obtained through 5,000 independent draws from the distribution of estimates.

protection resulting from reduced transmission of the vaccine-targeted pathogens<sup>22</sup>. As of 2018, an estimated 66.8% of age-eligible children in countries with PCV programmes in place received a full series of at least three PCV10/13 doses<sup>23</sup> (Supplementary Table 23). Under these coverage levels, we estimate that direct effects of PCV10/13 prevent 23.8 (4.2–52.0) million episodes of antibiotic-treated ARI annually among children aged 24–59 months in LMICs (Fig. 4 and Supplementary Tables 24, 25). This effect represents 42.4% (32.7–47.5%) of all antibiotic-treated ARI episodes that vaccine-serotype pneumococci would be expected to cause among children aged 24–59 months in LMICs, in the absence of PCV10/13 use. We estimated that an additional 21.7 (3.8–47.5) million episodes of antibiotic-treated ARI could be prevented by the direct effects of vaccination if PCV10/13 coverage were expanded to reach all children aged 24–59 months in LMICs. This residual burden corresponds to 38.7% (30.0–43.4%) of all antibiotic-treated ARI episodes that we estimated to be attributable to vaccine-serotype pneumococci among children aged 24–59 months in LMICs.

Roughly 77.3% of age-eligible children in LMICs in which rotavirus vaccination has been implemented received a full rotavirus vaccine series as of 2018<sup>23</sup>. We estimated that direct effects of rotavirus vaccination currently prevent 13.6 (3.6–23.7) million episodes of antibiotic-treated diarrhoea annually among children aged 0–23 months in LMICs (Supplementary Table 26). This effect reflects only 31.0% (17.7–35.2%) of all antibiotic-treated diarrhoea episodes that we would expect rotavirus to cause annually among children aged 0–23 months who live in LMICs, in the absence of vaccination. We estimated that an additional

18.3 (4.2–32.6) million episodes of antibiotic-treated diarrhoea could be prevented annually by the direct effects of rotavirus vaccination among children aged 0–23 months under a scenario of universal vaccine coverage, representing 42.1% (14.6% to 50.7%) of all antibiotic use that we estimated to be attributable to rotavirus in this population.

## Discussion

We found that vaccines against *S. pneumoniae* and rotavirus reduce antibiotic consumption among children in LMICs. In the absence of these vaccines, we estimated that PCV10/13-serotype pneumococci and rotavirus would cause 56.9 million and 47.9 million episodes of antibiotic-treated ARI and diarrhoea, respectively, among the age groups at greatest risk for these infections. Current PCV10/13 and rotavirus immunization programmes prevent an estimated 37.4 million episodes of antibiotic-treated illnesses annually through direct vaccine-conferred protection among children under five years of age in LMICs. Within this population, an additional 40.0 million episodes can be prevented through improvements in vaccine coverage in settings in which PCV10/13 and rotavirus vaccines are currently implemented, and through the introduction of these vaccines in countries in which children currently do not receive them.

It is helpful to consider the effects that we estimate alongside those of other interventions aiming to reduce unnecessary antibiotic consumption. Infrastructural improvements that address water, sanitation and hygiene have been prioritized in LMIC action plans to combat AMR<sup>8,9</sup>. However, randomized trials reporting suboptimal effects of such interventions on disease incidence<sup>24</sup> suggest that these approaches may present limited opportunities to prevent antibiotic-treated illness. Although certain antimicrobial stewardship interventions are feasible in LMICs<sup>25</sup>, the need to tailor such interventions to country-specific contexts hinders the development, adoption and assessment of universal strategies<sup>26</sup>. Recent aetiological studies reveal that most antibiotic courses administered to cases of ARI and diarrhoea among children in LMICs are unnecessary or inappropriate<sup>7,27</sup>, but few diagnostic tools exist to inform the clinical management of these infections. This shortage of alternative approaches to optimize antibiotic treatment in LMICs underscores the potential value of vaccination as a component of strategies to reduce AMR.

We used the recent introduction of PCV10/13 and rotavirus vaccines into paediatric immunization programmes as a natural means to estimate antibiotic use attributable to the vaccine-targeted pathogens. Strengths of our study include the opportunity to estimate the direct effects of the vaccines and disease burden using individual-level data from consistently designed surveys; statistical power for the estimation of small effect sizes in our case–control study of more than 60,000 children across 16 countries; and our ability to validate vaccine effectiveness estimates against negative-control end points. However, certain limitations should be considered. First, we relied on mother-reported disease outcomes and antibiotic use rather than a standardized clinical assessment. Although an imperfect data source<sup>28</sup>, maternal reports have been validated for the measurement of acute illness, care-seeking and antibiotic receipt among children in LMICs in previous studies<sup>15,29–31</sup>. Risk factor data in household surveys provided an opportunity to correct for potential differences in reporting thresholds for these conditions across settings—a limitation of previous burden studies based on survey data<sup>32,33</sup>. Second, our vaccine-analysis method permitted us to estimate pathogen-attributable antibiotic use only within the strata for which we had high-quality estimates of vaccine efficacy, as well as evidence of effect of the vaccines in our case–control study. For this reason, our findings centre on the age groups at greatest risk for pneumococcal ARI and rotavirus diarrhoea, among all children under five years of age who live in LMICs<sup>6,7</sup>. Finally, our focus on the direct effects of the vaccines, which we were able to estimate through a case–control study, may lead to underestimation of the population-level

impact of vaccination programmes. Indirect effects of PCV10/13 and rotavirus vaccines have also been reported in LMICs in which these vaccines have been implemented in routine paediatric immunization programmes<sup>34–37</sup>.

Several further considerations inform the interpretation of the effects of PCV10/13 on antibiotic consumption. Acute otitis media, sinusitis and other upper respiratory infections that are potentially attributable to pneumococci account for a substantial share of antibiotic use among infants within high-income settings<sup>38</sup>. These conditions may not be captured with high sensitivity by the ARI definition used in household surveys<sup>39,40</sup>. As a result, our use of an ARI end point is likely to underestimate the antibiotic consumption that is preventable by PCV10/13. By contrast, replacement of vaccine-targeted pneumococcal serotypes by non-vaccine serotypes may reduce the long-term effect of PCV10/13 on various pneumococcal disease end points<sup>41</sup>. However, the association of vaccine-targeted serotypes with conditions such as pneumonia and otitis media<sup>42,43</sup>, which drive antibiotic consumption, suggests a limited capacity of replacement serotypes to offset the effects of PCV10/13 on this consumption (Extended Data Fig. 1). In addition, the reported extent of vaccine-driven serotype replacement in pneumococcal carriage and disease has remained lower in LMICs<sup>35,44,45</sup> than in high-income settings, for reasons that remain incompletely understood<sup>46</sup>.

Emerging recognition of the threat posed by AMR has led to global, regional and national action plans to address the increasing burden of resistant disease<sup>8,47–49</sup>. A lack of evidence on the effectiveness of strategies advocated in these plans—which span public education efforts, human and veterinary antimicrobial stewardship, sanitation and other interventions—constrains the ability of policymakers to allocate resources to approaches with the greatest effect. Our study demonstrates that PCV10/13 and rotavirus vaccines reduce antibiotic consumption among children in LMICs, and that a substantial volume of antibiotic consumption may be prevented by introducing and increasing the global coverage of these vaccines. This evidence supports the prioritization of childhood vaccines as part of a global strategy to combat AMR.

## Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-020-2238-4>.

1. Klugman, K. P. & Black, S. Impact of existing vaccines in reducing antibiotic resistance: primary and secondary effects. *Proc. Natl Acad. Sci. USA* **115**, 12896–12901 (2018).
2. Lipsitch, M. & Siber, G. R. How can vaccines contribute to solving the antimicrobial resistance problem? *mBio* **7**, e00428-16 (2016).
3. Atkins, K. E. & Flasche, S. Vaccination to reduce antimicrobial resistance. *Lancet Glob. Health* **6**, e252 (2018).
4. Bloom, D. E., Black, S., Salisbury, D. & Rappuoli, R. Antimicrobial resistance and the role of vaccines. *Proc. Natl Acad. Sci. USA* **115**, 12868–12871 (2018).
5. Buckley, B. S. et al. Impact of vaccination on antibiotic usage: a systematic review and meta-analysis. *Clin. Microbiol. Infect.* **25**, 1213–1225 (2019).
6. Kotloff, K. L. et al. Burden and aetiology of diarrhoeal disease in infants and young children in developing countries (the Global Enteric Multicenter Study, GEMS): a prospective, case-control study. *Lancet* **382**, 209–222 (2013).
7. The Pneumonia Etiology Research for Child Health (PERCH) Study Group. Causes of severe pneumonia requiring hospital admission in children without HIV infection from Africa and Asia: the PERCH multi-country case-control study. *Lancet* **394**, 757–779 (2019).
8. World Health Organization. *Global Action Plan on Antimicrobial Resistance* (WHO Press, 2015).
9. The World Bank. *Drug-resistant Infections: A Threat to Our Economic Future*. <https://bit.ly/2x3mlqU> (2017).
10. UK Department of Health and Wellcome Trust. *The Global Economic Impact of Anti-microbial Resistance*. <https://assets.kpmg/content/dam/kpmg/pdf/2014/12/amr-report-final.pdf> (KPMG LLP, 2014).
11. Bell, B. G., Schellevis, F., Stobberingh, E., Goossens, H. & Pringle, M. A systematic review and meta-analysis of the effects of antibiotic consumption on antibiotic resistance. *BMC Infect. Dis.* **14**, 13 (2014).

12. Klein, E. Y. et al. Global increase and geographic convergence in antibiotic consumption between 2000 and 2015. *Proc. Natl Acad. Sci. USA* **115**, E3463–E3470 (2018).
13. Abat, C., Gautret, P. & Raoult, D. Benefits of antibiotics burden in low-income countries. *Proc. Natl Acad. Sci. USA* **115**, E8109–E8110 (2018).
14. National Vaccine Advisory Committee. A call for greater consideration for the role of vaccines in national strategies to combat antibiotic-resistant bacteria: recommendations from the National Vaccine Advisory Committee. *Public Health Rep.* **131**, 11–16 (2016).
15. Rogawski, E. T. et al. Use of antibiotics in children younger than two years in eight countries: a prospective cohort study. *Bull. World Health Organ.* **95**, 49–61 (2017).
16. Berman, S. Epidemiology of acute respiratory infections in children of developing countries. *Rev. Infect. Dis.* **13**, S454–S462 (1991).
17. International Vaccine Access Center (IVAC). *VIEW-hub Report: Global Vaccine Introduction and Implementation*. <https://view-hub.org/> (Johns Hopkins Bloomberg School of Public Health, 2018).
18. Berkley, S. Make vaccine coverage a key UN health indicator. *Nature* **526**, 165 (2015).
19. Pitzer, V. E. et al. Evaluating strategies to improve rotavirus vaccine impact during the second year of life in Malawi. *Sci. Transl. Med.* **11**, eaav6419 (2019).
20. Clark, A. et al. Efficacy of live oral rotavirus vaccines by duration of follow-up: a meta-regression of randomised controlled trials. *Lancet Infect. Dis.* **19**, 717–727 (2019).
21. Platts-Mills, J. A. et al. Pathogen-specific burdens of community diarrhoea in developing countries: a multisite birth cohort study (MAL-ED). *Lancet Glob. Health* **3**, e564–e575 (2015).
22. Halloran, M. E. & Struchiner, C. J. Causal inference in infectious diseases. *Epidemiology* **6**, 142–151 (1995).
23. WHO/UNICEF. *Immunization Coverage Estimates: 2018 Revision*. [https://www.who.int/immunization/monitoring\\_surveillance/routine/coverage/WUENIC\\_notes.pdf?ua=1](https://www.who.int/immunization/monitoring_surveillance/routine/coverage/WUENIC_notes.pdf?ua=1) (2019).
24. Cumming, O. et al. The implications of three major new trials for the effect of water, sanitation and hygiene on childhood diarrhea and stunting: a consensus statement. *BMC Med.* **17**, 173 (2019).
25. Van Dijk, C., Vlieghe, E. & Cox, J. A. Antibiotic stewardship interventions in hospitals in low-and middle-income countries: a systematic review. *Bull. World Health Organ.* **96**, 266–280 (2018).
26. Cox, J. A. et al. Antibiotic stewardship in low- and middle-income countries: the same but different? *Clin. Microbiol. Infect.* **23**, 812–818 (2017).
27. Platts-Mills, J. A. et al. Use of quantitative molecular diagnostic methods to assess the aetiology, burden, and clinical characteristics of diarrhoea in children in low-resource settings: a reanalysis of the MAL-ED cohort study. *Lancet Glob. Health* **6**, E1309–E1318 (2018).
28. Campbell, H. et al. Measuring coverage in MNCH: challenges in monitoring the proportion of young children with pneumonia who receive antibiotic treatment. *PLoS Med.* **10**, e1001421 (2013).
29. McCarthy, K. J. et al. Can surveys of women accurately track indicators of maternal and newborn care? A validity and reliability study in Kenya. *J. Glob. Health* **6**, 020502 (2016).
30. Carter, E. D. et al. Validity of maternal report of care-seeking for childhood illness. *J. Glob. Health* **8**, 010602 (2018).
31. Janmohamed, A. & Doledec, D. Comparison of administrative and survey data for estimating vitamin A supplementation and deworming coverage of children under five years of age in sub-Saharan Africa. *Trop. Med. Int. Health* **22**, 822–829 (2017).
32. Lanata, C. F. et al. Methodological and quality issues in epidemiological studies of acute lower respiratory infections in children in developing countries. *Int. J. Epidemiol.* **33**, 1362–1372 (2004).
33. Levine, G. A., Watson, J. L., Atlas, H. E., Lamberti, L. M. & Pavlinac, P. B. Defining pediatric diarrhea in low-resource settings. *J. Pediatr. Infect. Dis. Soc.* **6**, 289–293 (2017).
34. von Gottberg, A. et al. Effects of vaccination on invasive pneumococcal disease in South Africa. *N. Engl. J. Med.* **371**, 1889–1899 (2014).
35. Hammit, L. L. et al. Effect of ten-valent pneumococcal conjugate vaccine on invasive pneumococcal disease and nasopharyngeal carriage in Kenya: a longitudinal surveillance study. *Lancet* **393**, 2146–2154 (2019).
36. Ngabo, F. et al. Effect of pentavalent rotavirus vaccine introduction on hospital admissions for diarrhoea and rotavirus in children in Rwanda: a time-series analysis. *Lancet Glob. Health* **4**, E129–E136 (2016).
37. Bennett, A. et al. Direct and possible indirect effects of vaccination on rotavirus hospitalisations among children in Malawi four years after programmatic introduction. *Vaccine* **36**, 7142–7148 (2018).
38. Zhou, F., Shefer, A., Kong, Y. & Nuorti, J. P. Trends in acute otitis media-related health care utilization by privately insured young children in the United States, 1997–2004. *Pediatrics* **121**, 253–260 (2008).
39. Li, M. G., Hotez, P. J., Vrabec, J. T. & Donovan, D. T. Is chronic suppurative otitis media a neglected tropical disease? *PLoS Negl. Trop. Dis.* **9**, e0003485 (2015).
40. Monasta, L. et al. Burden of disease caused by otitis media: systematic review and global estimates. *PLoS ONE* **7**, e36226 (2012).
41. Weinberger, D. M., Malley, R. & Lipsitch, M. Serotype replacement in disease after pneumococcal vaccination. *Lancet* **378**, 1962–1973 (2011).
42. Lewnard, J. A., Givon-Lavi, N., Tähtinen, P. A. & Dagan, R. Pneumococcal phenotype and interaction with nontypeable *Haemophilus influenzae* as determinants of otitis media progression. *Infect. Immun.* **86**, e00727-17 (2018).
43. Faingelert, Y. et al. Nasopharyngeal carriage of invasive pneumococcal serotypes during childhood community-acquired alveolar pneumonia is associated with specific clinical presentation. *J. Infect. Dis.* **221**, 812–819 (2020).
44. Mackenzie, G. A. et al. Effect of the introduction of pneumococcal conjugate vaccination on invasive pneumococcal disease in The Gambia: a population-based surveillance study. *Lancet Infect. Dis.* **16**, 703–711 (2016).
45. Lourenço, J. et al. Determinants of high residual post-PCV13 pneumococcal vaccine-type carriage in Blantyre, Malawi: a modelling study. *BMC Med.* **17**, 219 (2019).
46. Lewnard, J. A. & Hanage, W. P. Making sense of differences in pneumococcal serotype replacement. *Lancet Infect. Dis.* **19**, e213–e220 (2019).

47. CDC. *U.S. National Action Plan for Combating Antibiotic-Resistant Bacteria (National Action Plan)*. <https://www.cdc.gov/drugresistance/us-activities/national-action-plan.html> (2015).
48. Essack, S. Y., Desta, A. T., Abotsi, R. E. & Agoba, E. E. Antimicrobial resistance in the WHO African region: current status and roadmap for action. *J. Publ. Health* **39**, 8–13 (2017).
49. HM Government. *Tackling Antimicrobial Resistance 2019–2024: The UK's Five-year National Action Plan*. [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/784894/UK\\_AMR\\_5\\_year\\_national\\_action\\_plan.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/784894/UK_AMR_5_year_national_action_plan.pdf) (2019).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020

# Article

## Methods

### Household survey data

We used data from Demographic Health Surveys (DHS; rounds V–VII; <https://www.dhsprogram.com>) and Multiple Indicator Cluster Surveys (MICS; rounds 5, 6; <http://mics.unicef.org/surveys>) undertaken in various LMICs from 2006 to 2018. Design of the two surveys was coordinated to facilitate data comparison and joint analyses across countries and over time. In brief, a two-stage probability sampling approach was used to select geographical clusters within countries or provinces of countries, and to select households within clusters. Survey questionnaires addressed household composition as well as risk factors, health outcomes and healthcare utilization among household occupants; original survey instruments can be viewed at <https://dhsprogram.com/publications/publication-DHSG4-DHS-Questionnaires-and-Manuals.cfm> (for DHS data) and <https://mics.unicef.org/surveys> (for MICS data).

Mothers were asked to answer survey questions on behalf of children under five years of age. Health information collected about children included history of ARI symptoms (defined as short, rapid breathing or difficult breathing that was chest-related or occurring with cough) and diarrhoea (defined as the occurrence of three loose stools in a 24-h period) in the 2 weeks preceding the survey. Mothers were asked about care seeking at public-sector or private-sector hospitals or clinics, pharmacies or doctor's offices, and receipt of drugs including antibiotic pills, syrups or injections, antimalarials, and analgesics or antipyretics. For diarrhoea, mothers were also asked about children's receipt of antimotility drugs, intravenous feeding, zinc and oral rehydration. Vaccination status (including antigens, doses and date of receipt) was collected from vaccination cards. Round VII of DHS and round 6 of MICS were the first to include PCV10/13 and rotavirus vaccination; thus, the case-control study (detailed below) included data from DHS round VII and MICS round 6 only.

### Case-control study

We estimated PCV10/13 effectiveness against ARI-related end points and rotavirus vaccine effectiveness against diarrhoea-related end points in an individually matched, case-control study to compare counterfactual outcomes among individuals encountering similar conditions of transmission, as well as other exposures (the matching procedure is described below). End points were: occurrence of ARI and diarrhoea symptoms; occurrence of ARI and diarrhoea symptoms for which medical treatment or advice was sought (from a source other than a traditional practitioner); and occurrence of ARI and diarrhoea symptoms for which antibiotic pills, syrups or injections were received. For diarrhoea analyses, we also considered an end point of diarrhoea symptoms resulting in oral rehydration using pre-packaged salts or fluid, or recommended homemade fluids or intravenous solution, among children who sought care (results are included in Supplementary Tables 4, 5). For the purposes of our analyses, children experiencing each end point in the preceding 2 weeks were defined as 'cases', and those who did not report symptoms were defined as controls.

Within each country, we matched cases and controls on the basis of age and visit timing (each within one month), within-country wealth quintile (a variable defined on the basis of country-specific household socioeconomic indicators), urbanicity (residence within a sampling cluster defined as urban or rural) and pentavalent vaccine doses received (to reduce confounding by healthcare access, including access to vaccination, and to control for any effects of diphtheria, pertussis and *Haemophilus influenzae* type b antigens on ARI). Healthcare utilization questions for the ARI end point did not distinguish whether care was received due to respiratory symptoms or due to fever. Therefore, to separate vaccine effects from antibiotic treatment associated with ARI symptoms (rather than treatment of fever cases that would be expected

to occur independent of ARI symptoms, for example, due to malaria, typhoid or arboviruses), we also matched children on mother-reported fever and excluded those receiving antimalarials.

We matched cases and controls without replacement at a 1:3 ratio and undertook statistical inference in a resampling framework, generating 2,000 independent draws from the permutation distribution of case-control match assignments. Within each iteration, we estimated the matched odds ratio ( $OR_M$ ) using conditional logistic regression. Defining  $Y = 1$  and  $Y = 0$  as case and control status, respectively, and  $V = 1$  and  $V = 0$  as vaccinated and unvaccinated status, respectively, we took  $1 - OR_M(V|Y)$  to measure vaccine effectiveness.

### Assessment of residual confounding

The relatively recent implementation of PCV10/13 and rotavirus vaccines in the immunization programmes of countries may make pentavalent vaccine an imperfect surrogate of the access of children to these newer products. In particular, within-country geographical differences in access to PCV10/13 and rotavirus vaccines may introduce residual confounding if the differences in vaccine access are associated with geographical variation in disease burden.

Data from DHS and MICS include the subnational administrative regions in which children reside, although this variable must be interpreted with caution as the size and uniformity of subnational regions varies across countries (between 4 and 3,399 children were available from each region (median, 217 children), before accounting for match eligibility). Although it was not possible to match children at precise subnational resolutions for our case-control study, we used subnational administrative region designators to assess risk for residual confounding by geography. We computed the proportion of variance in ARI and diarrhoea outcomes, and in PCV10/13 and rotavirus vaccine receipt, explained by subnational region after accounting for all other matching factors included in our analysis. Specifically, we assessed the difference in the summed squared error of residuals from regression models treating disease end points and vaccination as outcome variables, which either included or did not include subnational administrative units as intercepts after controlling for age and visit timing (each as monthly factor variables), urbanicity, wealth quintile, mother-reported fever (only for analyses of ARI and PCV10/13) and pentavalent vaccine doses received. Low estimates of the proportion of variance explained by subnational region in either the exposure or outcome suggest minimal risk of confounding by geography. The results of this analysis are included in Supplementary Table 6. In addition, to assess whether negative-control analyses provided a sufficient opportunity to correct for residual confounding associated with access to these newer vaccines, we estimated the correlation in receipt of full PCV10/13 and rotavirus vaccine series among age-eligible children in countries where both vaccines were recommended. We describe the results of these analyses in the Supplementary Information section 'Assessment of residual confounding'.

### Vaccine probe study

We formulated a simple model that enables the estimation of the proportion of episodes attributable to vaccine-targeted pathogens for each end point (for example, ARI symptoms or antibiotic-treated ARI symptoms) in the absence of vaccination. We defined  $\rho$  as the risk for a child to experience a given end point attributable to the vaccine-targeted pathogen in the 2-week recall period,  $\omega$  as the risk of the same end point attributable to all other causes and  $\theta$  as the relative risk of disease caused by the vaccine-targeted pathogen in a vaccinated individual relative to an unvaccinated individual, owing only to vaccine-conferred protection (such that  $1 - \theta$  measures vaccine efficacy against illness caused by the targeted pathogen). Under this framework



$$\begin{aligned} \text{OR}_M(V|Y) &= \frac{\Pr(V=1, Y=1)\Pr(V=0, Y=0)}{\Pr(V=0, Y=1)\Pr(V=1, Y=0)} \\ &= \frac{[\theta\rho + \omega][1 - \rho - \omega]}{[1 - \theta\rho - \omega][\rho + \omega]} \\ &= \frac{\theta\rho + \omega}{\rho + \omega} \times \frac{1 - \rho - \omega}{1 - \theta\rho - \omega} \end{aligned} \quad (1)$$

Noting that  $\frac{1 - \rho - \omega}{1 - \theta\rho - \omega} \approx 1$  for small  $\rho$  and  $\omega$

$$\text{OR}_M(V|Y) \approx \frac{\theta\rho + \omega}{\rho + \omega} \quad (2)$$

or in other words, we expected the matched odds ratio to provide a reasonable estimator for the relative hazard of each end point, given vaccination. Using equation (2), above

$$\rho \approx \frac{\omega(1 - \text{OR}_M(V|Y))}{\text{OR}_M(V|Y) - \theta} \quad (3a)$$

and

$$\frac{\rho}{\rho + \omega} \approx \frac{1 - \text{OR}_M(V|Y)}{1 - \theta} \quad (3b)$$

The above approximation relies on the assumption that  $\text{OR}_M(V|Y)$  is similar to the relative risk of disease, given vaccination. We illustrate in Extended Data Fig. 2 the potential degree of bias under differing values of  $\theta$ ,  $\rho$  and  $\omega$  that are concordant with our results and those of external studies assessing all-cause burden of ARI and diarrhoea among children in LMICs<sup>15,50,51</sup>. The degree of bias approaches zero under any of the following conditions: (i) the all-cause disease end point is rare (that is,  $\rho + \omega \approx 0$ ); (ii) the attributable fraction is low ( $\rho \ll \omega$ ); or (iii) the vaccine is highly efficacious against disease due to the pathogen of interest ( $\theta \approx 0$ ).

### Estimation of the fraction attributable to rotavirus

Because previous studies have reported differential efficacy and effectiveness of rotavirus vaccines in socioeconomically distinct settings, we generated stratified estimates of  $\text{OR}_M(V|Y)$  (in the case-control study) and  $\theta$  (in the meta-analysis, as described below) for children aged 0–23 months residing in middle-income and low-income countries. We used these inputs to generate separate rotavirus-attributable fraction estimates associated with each end point for middle-income and low-income countries. We reconstructed the distribution of the rotavirus-attributable fraction for each end point across all LMIC settings (Fig. 2 and Supplementary Table 8) using the weighted average of paired draws from the independent distributions of the attributable fraction estimates of middle-income and low-income countries; weights were proportional to the total populations of children aged 0–23 months residing in middle-income and low-income countries.

To account for differential direct effects of the vaccine against rotavirus-attributable diarrhoea end points in middle-income and low-income countries, we used our stratified estimates of rotavirus vaccine efficacy, effectiveness and attributable fractions for all burden assessments.

### Meta-analysis of vaccine efficacy studies

We obtained estimates of  $\theta$  for PCV10/13 and rotavirus vaccine using a meta-analysis of vaccine efficacy studies. Because  $\theta$  has not been estimated for the effects of PCV10/13 against vaccine-serotype ARI, we took PCV efficacy against invasive disease caused by vaccine-serotype pneumococci as a primary estimate of protection, and PCV efficacy against acute otitis media caused by vaccine-serotype pneumococci as

a lower-bound estimate, consistent with previous studies<sup>52</sup>. Estimates of PCV efficacy against these end points were aggregated from two previously published reviews<sup>53,54</sup>. We used estimates of efficacy against invasive pneumococcal disease from only low- and middle-income countries. Because such estimates were not available for PCV efficacy against acute otitis media, we did not restrict estimates for this end point on the basis of study setting.

For rotavirus, we used estimates of  $\theta$  from studies that estimated the efficacy of the human pentavalent rotavirus vaccine (Rotarix) against rotavirus gastroenteritis in the first two years of life in LMICs, again aggregated in a previous systematic review<sup>55</sup>. Data were restricted to Rotarix studies because this was the only vaccine used in the national immunization programmes of countries included in the case-control study.

For both end points, we estimated pooled values of  $\theta$  using inverse variance-weighted random effect models using log-transformed effect size estimates from each study.

### Risk factor analysis of household survey data

We used a regression-based approach to estimate country-level rates of incidence of ARI and diarrhoea in the year 2016 in the absence of PCV10/13 and rotavirus vaccine use. Because previous studies have described marked variation in the clinical threshold for maternal or caregiver report of ARI and diarrhoea<sup>28,32,33,56</sup>, we aimed to reconstruct incidence rates on the basis of risk factor prevalence, controlling for region-level intercepts that arise from differential reporting of symptoms. We built Poisson regression models for the outcome of reporting of syndrome  $Z$  (signifying all-cause ARI or diarrhoea) for child  $i$  of the form

$$E(Z_i|X_i) = \pi_{R(i)} \times \exp\left(\alpha + \sum_k \beta_k X_{k,i} + \epsilon_i\right) \quad (4)$$

where  $\alpha$  represents a general intercept for the log-transformed incidence rate. Each variable  $X_{k,i}$  indicates the exposure of child  $i$  to the  $k$ th risk factor for syndrome  $Z$ , associated with a  $\beta_k$ -fold increase in incidence; the  $\epsilon_i$  errors are independent and identically distributed with mean zero among all children. The term  $\pi_{R(i)}$  indicates the probability of reporting syndrome  $Z$  given it truly occurred among children in the same region as child  $i$ . We extracted the risk factors listed in Supplementary Table 27 on the basis of their inclusion in both DHS and MICS datasets, and based on previous evidence of their relevance to the risk of ARI and/or diarrhoea for children<sup>51,57,58</sup>. In addition to variables included in DHS and MICS, we extracted the gross domestic product (GDP) per capita of each country in the survey year and in 2016. We included covariates in the model identified to have a statistically significant ( $P < 0.05$ ) unadjusted association with the outcome variable that persisted in multivariable analyses, and tested for significant pairwise interactions among covariates. Risk factor estimates are shown in Extended Data Fig. 3 and Supplementary Tables 28–30 for ARI (for children aged 24–59 months and 0–59 months) and diarrhoea (for children aged 0–23 months).

We similarly modelled the probability ( $H$ ) for ARI and diarrhoea to be treated with antibiotics using a Poisson regression model of the form

$$E(H_i|X_i) = \exp\left(\delta + \sum_k \delta_k X_{k,i} + \eta_i\right) \quad (5)$$

We evaluated the same risk factors for inclusion as in our analyses of ARI and diarrhoea; here we defined  $\eta_i$  as mean-zero independent and identically distributed error terms. The individual risk factor estimates are shown in Extended Data Figs. 4, 5 and Supplementary Tables 31–33 for antibiotic-treated ARI and diarrhoea.

We limited analyses of ARI and diarrhoea end points to children who had received no PCV10/13 doses and no rotavirus vaccine doses,

respectively. For DHS rounds V–VI and MICS round 5, which did not include PCV10/13 and rotavirus vaccine data collection, children were assumed not to have received PCV10/13 or rotavirus vaccine doses if countries had not yet implemented these vaccines in their immunization programmes at the time of the survey. Surveys undertaken in countries that had implemented PCV10/13 or rotavirus vaccines were excluded if information on the receipt of these vaccines among children was not collected.

Missing values in the outcome and risk factor variables were populated in five independent pseudo-datasets by multiple imputation using the Amelia II package in R<sup>59</sup>; in total, 11.4% of data cells were missing at the outset of analyses, including 0.6% and 2.4% of ARI and antibiotic-treated ARI observations, and 0.1% and 0.4% of diarrhoea and antibiotic-treated diarrhoea observations. Additional measures<sup>12</sup> of country-level antibiotic access (Supplementary Table 27) were included due to scientific interest and a consideration that these covariates may aid imputation.

## Assumption of a Poisson outcome distribution

Translating the reported history of any ARI or diarrhoea in the preceding 2 weeks to a Poisson-distributed number of ARI or diarrhoea episodes required two assumptions: that episodes are acute (consistent with the nature of illness solicited by DHS and MICS household surveys) and that overall incidence rates are low (resulting in a low or negligible likelihood of  $\geq 2$  distinct episodes occurring in a single fortnightly period in which a child was reported to have experienced ARI or diarrhoea)<sup>30</sup>. The low frequency of diarrhoea lasting 14 days or longer in a previous cohort study in LMICs<sup>21</sup> (accounting for 4.9% and 1.8% of all diarrhoea episodes in the first and second years of life, respectively) indicated that our assumption of an acute nature of mother-reported illness was appropriate for the diarrhoea outcome. Similarly, previous studies of respiratory infections among children in LMICs have reported short average durations of new-onset disease episodes, with 87% to 100% of episodes lasting  $\leq 2$  weeks<sup>60–62</sup>. For incidence rates similar to those estimated in our study (approximately 4 diarrhoea cases and 1.25 ARI cases per child per year), assuming a 7-day average duration of illness and a definition of 3 symptom-free days to distinguish separate disease episodes, the assumption of exponentially distributed inter-event times (concordant with Poisson-distributed counts) would yield a second new-onset diarrhoea case in only 4.3% of 2-week reporting periods with a first case identified, and a second new-onset ARI case in only 1.4% of reporting periods with a first case identified. This small bias, if applicable to our analyses, would be expected to lead to slight underestimation of true burden.

## Incidence rate estimation

From the fitted models, we generated standardized estimates of the hazard rates of ARI and diarrhoea for individual children for the year 2016. To correct for differential reporting, we took the lowest and highest estimated reporting probability from any region ( $\pi_R$ ) to provide lower and upper bounds, respectively, on true incidence rates.

We sampled from the multivariate normal distribution of (log-transformed) regression parameter estimates and model residuals, thus accounting for uncertainty in the effects of risk factors on event rates. To propagate uncertainty in our analyses driven by sampling variability in the surveyed population as well as our multiple imputation procedure, we repeated this analysis 1,000 times for each of 5 multiply imputed datasets. At each iteration, we resampled children according to survey weights before sampling ARI and diarrhoea outcomes stochastically based on individual-level risk factors. On the basis of the estimated incidence rate of ARI and diarrhoea of each child from these risk factors, we sampled a total number of annual cases assuming this value would be Poisson-distributed with respect to the underlying individual-specific rate. We used the same approach to sample the antibiotic treatment status of each case based on model-generated

estimates of the probability of antibiotic treatment, given individual risk factors, under an assumed Bernoulli outcome distribution.

## Extrapolation of burden estimates to countries without household survey data

For the 58 LMICs not covered by DHS rounds V–VII or MICS rounds 5 and 6, we extrapolated rates of incidence of all cases and antibiotic-treated cases of ARI and diarrhoea on the basis of national-level variables aggregated as Health, Nutrition and Population Statistics by the World Bank (<https://databank.worldbank.org/source/health-nutrition-and-population-statistics>). For each of 5,000 draws from the distribution of country-specific incidence-rate and treatment probability estimates (1,000 draws each from 5 multiply imputed datasets), we fitted regression tree models using tenfold cross-validation to a randomly sampled 90% set of countries via stochastic gradient boosting using the caret package in R with default tuning parameters<sup>63</sup>. We saved predictions for the 10% ‘holdout’ set of countries to assess model performance (Extended Data Fig. 6). We resampled the 90/10 set of training and holdout countries at each iteration. We used fitted models to generate out-of-sample predictions for countries without DHS and MICS data.

## Estimating the vaccine-preventable disease burden

We multiplied independent draws from the distribution of our estimates of the proportion of antibiotic-treated ARI episodes and antibiotic-treated diarrhoea episodes attributable to vaccine-serotype pneumococci and rotavirus, respectively, by independent draws from the distribution of our estimates of country-level incidence rates for antibiotic-treated ARI (for children aged 24–59 months and 0–59 months) and antibiotic-treated diarrhoea (for children aged 0–23 months). We used our stratified estimates of rotavirus-attributable fractions (for low-income and middle-income countries) for all analyses of diarrhoea burden end points. For analyses estimating vaccine direct effects, we multiplied independent draws from the distribution of our estimates of PCV10/13 direct effects against antibiotic-treated ARI and rotavirus vaccine direct effects against antibiotic-treated diarrhoea, as obtained in the case-control study, by independent draws from the distribution of our estimates of country-level incidence rates for antibiotic-treated ARI (of children aged 24–59 months and 0–59 months) and antibiotic-treated diarrhoea (in children aged 0–23 months).

To generate estimates of burden across multiple countries (globally, or grouped by income strata), we estimated total cases in each country by multiplying independent draws of the incidence rate for each end point by the number of children in each age group (using World Bank estimates; <https://databank.worldbank.org/home>). For each sampled parameterization (defined by estimates of vaccine effectiveness, pathogen-attributable burden and incidence rates), we summed total cases across countries belonging to a stratum of interest and estimated incidence by dividing total summed cases by the population at risk.

## Potential reduction in PCV10/13 effects owing to serotype replacement

Replacement of vaccine-targeted pneumococcal serotypes by nonvaccine serotypes may partially offset the estimated effect of PCV10/13 on all-cause ARI end points. We assessed the extent to which this may alter our estimates by determining the maximal increase in nonvaccine-type antibiotic-treated ARI that could be expected under scenarios consistent with reported serotype replacement in pneumococcal carriage in LMIC settings.

We defined  $p_{VT}(t)$  and  $p_{NVT}(t)$  as the prevalence of carriage of vaccine-type (VT) pneumococci and nonvaccine-type (NVT) pneumococci at time  $t$ , and  $r_{VT}$  and  $r_{NVT}$  as the rates of progression of vaccine-type and nonvaccine-type pneumococci from carriage to ARI<sup>42</sup>. We used

previous estimates of the relative incidence of vaccine-type and nonvaccine-type invasive pneumococcal disease per carrier<sup>64</sup>, and of vaccine-type and nonvaccine-type otitis media per carrier<sup>65</sup>, to supply bounds on  $r_{VT}$  and  $r_{NVT}$  for ARI, as this quantity has not previously been measured. We used data from three studies of pneumococcal carriage among children less than five years old before and after vaccine introduction in settings with long-term, continuous, prospective surveillance in place<sup>7,44,66</sup> to supply pre-vaccination ( $p_{VT}(0)$  and  $p_{NVT}(0)$ ) and post-vaccination ( $p_{VT}(1)$  and  $p_{NVT}(1)$ ) estimates of carriage prevalence. We took  $[p_{NVT}(1) - p_{NVT}(0)]r_{NVT}$  to indicate the excess incidence of disease attributable to nonvaccine serotypes that would result from post-vaccination replacement of vaccine-targeted serotypes by non-vaccine serotypes.

We compared this excess post-vaccination incidence attributable to serotype replacement with two measures of pre-vaccination disease incidence. The ratio  $[p_{NVT}(1) - p_{NVT}(0)]r_{NVT}/[p_{VT}(0)r_{VT}]$  compared the excess replacement-associated incidence of nonvaccine-serotype disease to the incidence of vaccine-serotype disease targetable by vaccination with PCV10/13. Second, defining AF as the fraction of all-cause disease incidence attributable to vaccine-type pneumococci, the ratio  $[p_{NVT}(1) - p_{NVT}(0)]r_{NVT}AF/[p_{VT}(0)r_{VT}]$  compared the extent of replacement-driven nonvaccine-serotype disease incidence to all-cause incidence in the pre-vaccination era. These estimates are shown in Extended Data Fig. 1, using the attributable fraction estimates for all-cause antibiotic-treated ARI (as this was the primary end point of interest in our analyses).

In light of the assumptions informing our interpretation of PCV10/13 effectiveness against non-specific ARI end points in the case-control study, we consider the estimates provided by this analysis to represent an upper bound on the potential replacement-driven disease burden. Our estimates of vaccine-serotype pneumococcal attributable fractions based on the case-control study assumed no effect of PCV10/13 on disease not attributable to vaccine-serotype pneumococci. However, children receiving PCV10/13 in prelicensure or early-implementation studies in fact have been reported to experience higher risk of carriage of nonvaccine serotypes<sup>67</sup> and resulting nonvaccine-type disease<sup>68,69</sup>. As such, odds ratio estimates in our case-control study may be interpretable as representing the ‘net’ reduction in disease risk among vaccinated children, accounting for both protection against vaccine-type pneumococci and increased acquisition of nonvaccine types.

## Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

## Data availability

MICS data are publicly available at <http://mics.unicef.org/>. DHS data are available upon request from <https://dhsprogram.com/Data/>. Files posted in the GitHub repository (<http://github.com/joelewnard>) include the reduced versions of the DHS and MICS datasets necessary to replicate analyses.

## Code availability

The analysis code necessary to replicate all parts of the analysis and for the generation of the figures is available from GitHub (<http://github.com/joelewnard>).

50. Fischer Walker, C. L., Perin, J., Aryee, M. J., Boschi-Pinto, C. & Black, R. E. Diarrhea incidence in low- and middle-income countries in 1990 and 2010: a systematic review. *BMC Publ. Health* **12**, 220 (2012).
51. Shi, T. et al. Global, regional, and national disease burden estimates of acute lower respiratory infections due to respiratory syncytial virus in young children in 2015: a systematic review and modelling study. *Lancet* **390**, 946–958 (2017).
52. Wahl, B. et al. Burden of *Streptococcus pneumoniae* and *Haemophilus influenzae* type b disease in children in the era of conjugate vaccines: global, regional, and national estimates for 2000–15. *Lancet Glob. Health* **6**, E744–E757 (2018).
53. Lucero, M. G. et al. Pneumococcal conjugate vaccines for preventing vaccine-type invasive pneumococcal disease and X-ray defined pneumonia in children less than two years of age. *Cochrane Database Syst. Rev.* **4**, CD004977 (2009).
54. Fortanier, A. C. et al. Pneumococcal conjugate vaccines for preventing otitis media. *Cochrane Database Syst. Rev.* **4**, CD001480 (2014).
55. Lamberti, L. M., Ashraf, S., Walker, C. L. F. & Black, R. E. A systematic review of the effect of rotavirus vaccination on diarrhea outcomes among children younger than 5 years. *Pediatr. Infect. Dis. J.* **35**, 992–998 (2016).
56. Hancioglu, A. & Arnold, F. Measuring coverage in MNCH: tracking progress in health for women and children using DHS and MICS household surveys. *PLoS Med.* **10**, e1001391 (2013).
57. GBD Diarrhoeal Diseases Collaborators. Estimates of global, regional, and national morbidity, mortality, and aetiologies of diarrhoeal diseases: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet Infect. Dis.* **17**, 909–948 (2017).
58. GBD 2017 Influenza Collaborators. Mortality, morbidity, and hospitalisations due to influenza lower respiratory tract infections, 2017: an analysis for the Global Burden of Disease Study 2017. *Lancet Respir. Med.* **7**, 69–89 (2019).
59. Honaker, J., King, G. & Blackwell, M. Amelia II: a program for missing data. *J. Stat. Softw.* **45**, 1–47 (2011).
60. Suttmoller, F. & Maia, P. R. Acute respiratory infections in children living in two low income communities of Rio de Janeiro, Brazil. *Mem. Inst. Oswaldo Cruz* **90**, 665–674 (1995).
61. Sharma, S. et al. Indoor air quality and acute lower respiratory infection in Indian urban slums. *Environ. Health Perspect.* **106**, 291–297 (1998).
62. Peasah, S. K. et al. The cost of acute respiratory infections in Northern India: a multi-site study. *BMC Public Health* **15**, 330 (2015).
63. Kuhn, M. Building predictive models in R using the caret package. *J. Stat. Softw.* **28**, 1–26 (2008).
64. Shouval, D. S., Greenberg, D., Givon-Lavi, N., Porat, N. & Dagan, R. Serotype coverage of invasive and mucosal pneumococcal disease in Israeli children younger than 3 years by various pneumococcal conjugate vaccines. *Pediatr. Infect. Dis. J.* **28**, 277–282 (2009).
65. Lewnard, J. A., Givon-Lavi, N., Weinberger, D. M., Lipsitch, M. & Dagan, R. Pan-serotype reduction in progression of *Streptococcus pneumoniae* to otitis media after rollout of pneumococcal conjugate vaccines. *Clin. Infect. Dis.* **65**, 1853–1861 (2017).
66. Heinsbroek, E. et al. Pneumococcal carriage in households in Karonga District, Malawi, before and after introduction of 13-valent pneumococcal conjugate vaccination. *Vaccine* **36**, 7369–7376 (2018).
67. Lewnard, J. A. et al. Epidemiological markers for interactions among *Streptococcus pneumoniae*, *Haemophilus influenzae*, and *Staphylococcus aureus* in upper respiratory tract carriage. *J. Infect. Dis.* **213**, 1596–1605 (2016).
68. Whitney, C. G. et al. Effectiveness of seven-valent pneumococcal conjugate vaccine against invasive pneumococcal disease: a matched case-control study. *Lancet* **368**, 1495–1502 (2006).
69. Eskola, J. et al. Efficacy of a pneumococcal conjugate vaccine against acute otitis media. *N. Engl. J. Med.* **344**, 403–409 (2001).
70. Usuf, E. et al. Persistence of nasopharyngeal pneumococcal vaccine serotypes and increase of nonvaccine serotypes among vaccinated infants and their mothers 5 years after introduction of pneumococcal conjugate vaccine 13 in The Gambia. *Clin. Infect. Dis.* **68**, 1512–1521 (2019).

**Acknowledgements** We thank P. Srikanthiah, J. Vekemans, M. Hasso-Agopsowicz, H. Prudden, E. Tayler, L. Simonsen, M. Lipsitch and members of the WHO VAC-AMR working group for comments. This work was funded by the Bill & Melinda Gates Foundation (grant number OPP1190803).

**Author contributions** Conceptualization: J.A.L. and R.L.; data curation, formal analysis, investigation, visualization, writing (original draft): J.A.L.; funding acquisition: J.A.L., N.A., I.F. and R.L.; project administration: J.A.L., I.F. and R.L.; methodology: J.A.L. and N.C.L.; writing (review and editing): J.A.L., N.C.L., N.A., I.F. and R.L.

**Competing interests** J.A.L. has received consulting fees and research grants from Pfizer, consulting fees and research grants from Merck Sharp & Dohme and research grants from the World Health Organization, all for unrelated work. N.C.L. has received personal fees from the World Health Organization, for unrelated work. All other authors declare no competing interests.

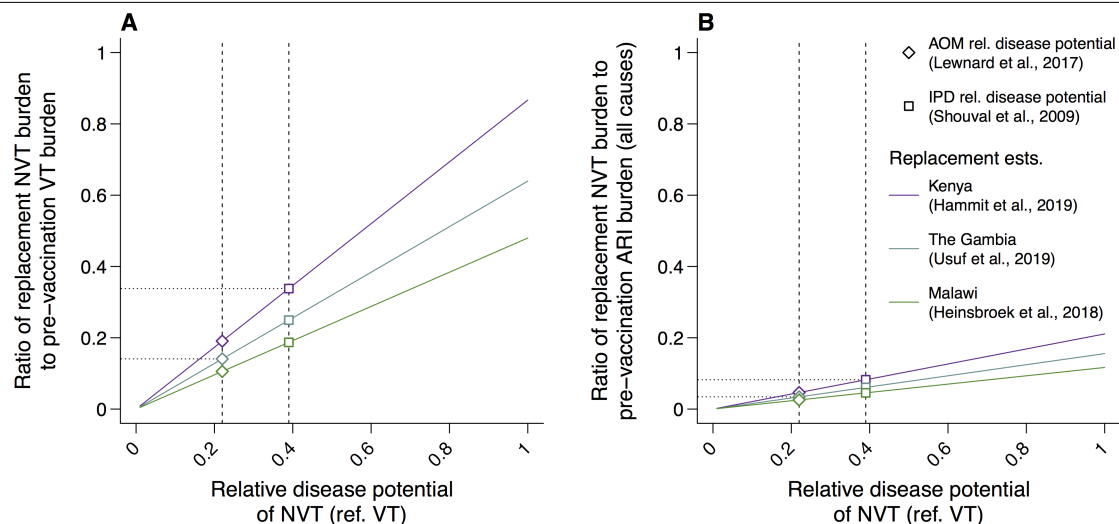
## Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41586-020-2238-4>.

**Correspondence and requests for materials** should be addressed to J.A.L.

**Peer review information** Nature thanks Uri Obolski and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

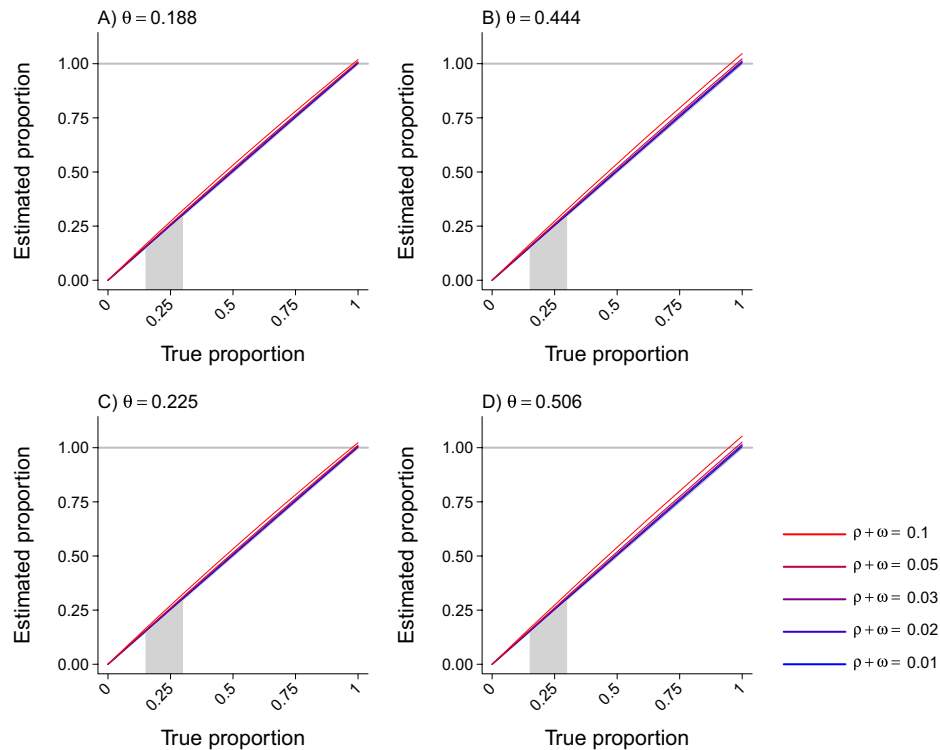
**Reprints and permissions information** is available at <http://www.nature.com/reprints>.



**Extended Data Fig. 1 | Potential effect of pneumococcal serotype replacement.** **a**, We compare the maximum estimate of the replacement disease associated with increased carriage of nonvaccine-type (NVT) pneumococci to the prevented burden of disease associated with vaccine-type (VT) pneumococci based on the approach described in the Methods section ‘Potential reduction in PCV10/13 effects owing to serotype replacement’. We plot the ratio of replacement-attributable NVT disease as a function of the relative pathogenicity of NVT and VT pneumococci. As this ratio may differ according to disease end point, the measure presented in **a** can be interpreted as end point agnostic. Estimates from two previous studies<sup>64,65</sup> of acute otitis media (AOM) and invasive pneumococcal disease (IPD) provide a range of 0.218–0.387 for the relative pathogenicity of NVT compared with VT pneumococci in these conditions. We plot estimates based on serotype

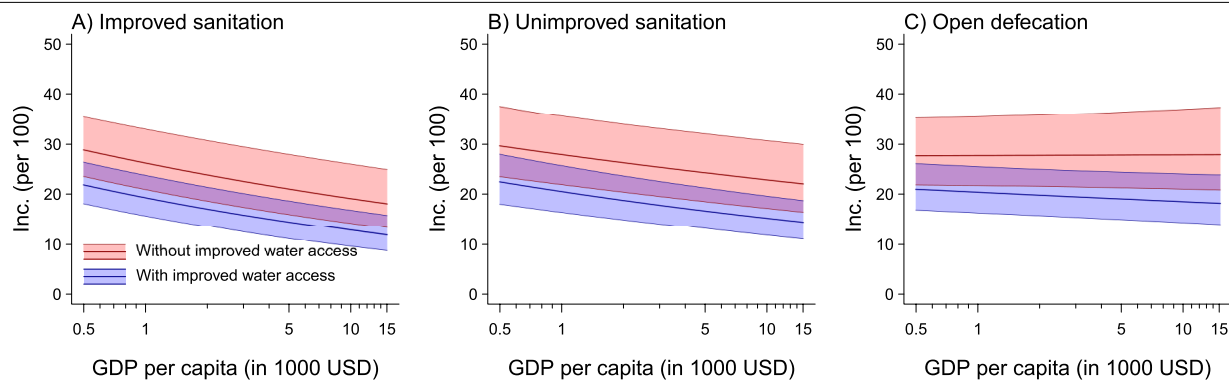
replacement observations from three carriage studies in LMICs with continuous, prospective surveillance in place before and after PCV10/13 implementation<sup>35,66,70</sup>. **b**, We next illustrate the estimated ratio of replacement-attributable antibiotic-treated ARI to all-cause antibiotic-treated ARI among children aged 24–59 months, based on our estimates of the fraction of antibiotic-treated ARI attributable to VT pneumococci (Fig. 2). On the basis of the input values from the carriage and disease studies cited<sup>35,64–66,70</sup>, we infer that the maximum extent of antibiotic-treated ARI that replacement serotypes would account for would decrease to between 2.6% and 8.2% of the pre-vaccination incidence of all-cause antibiotic-treated ARI; this range is well below the estimated reduction associated with protection against VT pneumococci.





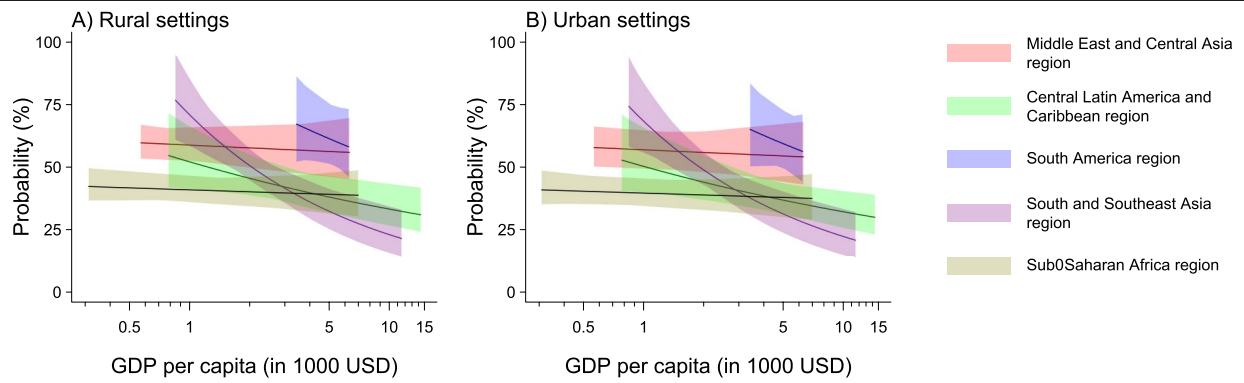
**Extended Data Fig. 2 | Bias that occurs when using the odds ratio to approximate the relative risk. a–d.** We illustrate the degree of bias that occurs in attributable fraction estimates (indicated by the departure of the estimated proportion from the 1:1 diagonal) under differing parameterizations with respect to the true aetiological fraction and vaccine efficacy against the targeted infection. Values correspond to our meta-analytic estimates of PCV efficacy against vaccine-serotype invasive pneumococcal disease (**a**), PCV efficacy against vaccine-serotype acute otitis media (**b**), rotavirus vaccine efficacy against rotavirus gastroenteritis in middle-income countries (**c**) and rotavirus vaccine efficacy against rotavirus gastroenteritis in low-income countries (**d**). The range of 15–30% is highlighted in grey as plausible values for

the proportion of disease attributable to vaccine-serotype pneumococci for children aged 24–59 months and rotavirus for children aged 0–23 months, based on previously published studies<sup>6,7</sup>. Values are plotted on a blue-to-red ramp corresponding to increases in symptom prevalence ( $p + \omega$ ); values of 0.01, 0.02, 0.03, 0.05 and 0.1 correspond to incidence rates of 52, 104, 156, 261 and 521 episodes per 100 children annually (roughly the range of our all-cause ARI and diarrhoea incidence rate estimates); under the assumption of a 3-day duration of symptoms, the same prevalence values correspond to incidence rates of 122 to 1,217 episodes per 100 children annually. These outcomes suggest negligible bias in aetiological fractions over the range of plausible values for our analysis.



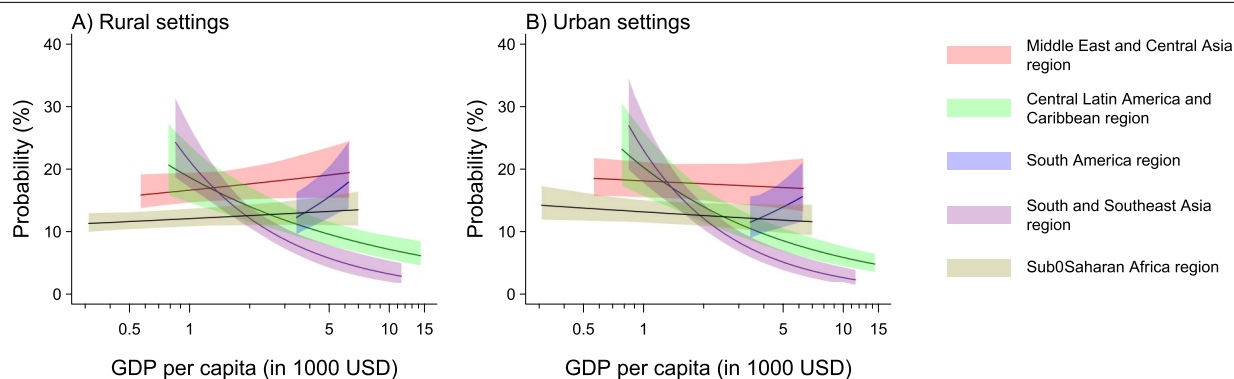
**Extended Data Fig. 3 | Fitted association of improved water and sanitation access with diarrhoea risk. a–c,** We illustrate model-estimated rates of the incidence (Inc.) per 100 children of diarrhoea under differing conditions of access to improved water and sanitation conditions among at ages 0–23 months. Estimates account for interactions of sanitation and water access with GDP per capita. Shaded regions indicate 95% confidence intervals around estimates. True intercepts are dependent on the distribution of other individual and setting-level risk factors; plotted estimates do not account for

the joint distribution of other risk factors with respect to the access of children to water and sanitation at the country level and GDP per capita. Sanitation conditions include improved (a) and unimproved sanitation (b) or open defaecation (c) in combination with or without improved water access. Model parameter estimates are presented in full in Supplementary Table 30. Estimates are obtained from analyses of DHS and MICS surveys comprising 377,665 children across 77 countries. Quantiles are obtained through 5,000 independent draws from the distribution of estimates.



**Extended Data Fig. 4 | Fitted association of GDP per capita with antibiotic treatment of ARI, by region. a, b,** We illustrate model-estimated probabilities of antibiotic treatment for cases of ARI among children aged 24–59 months in rural (**a**) and urban (**b**) settings associated with region and GDP per capita. Estimates for central and eastern Europe are omitted from plots as wide uncertainty intervals would obscure the illustration of associations in other settings. Shaded regions indicate 95% confidence intervals around the estimates. True intercepts are dependent on the distribution of other

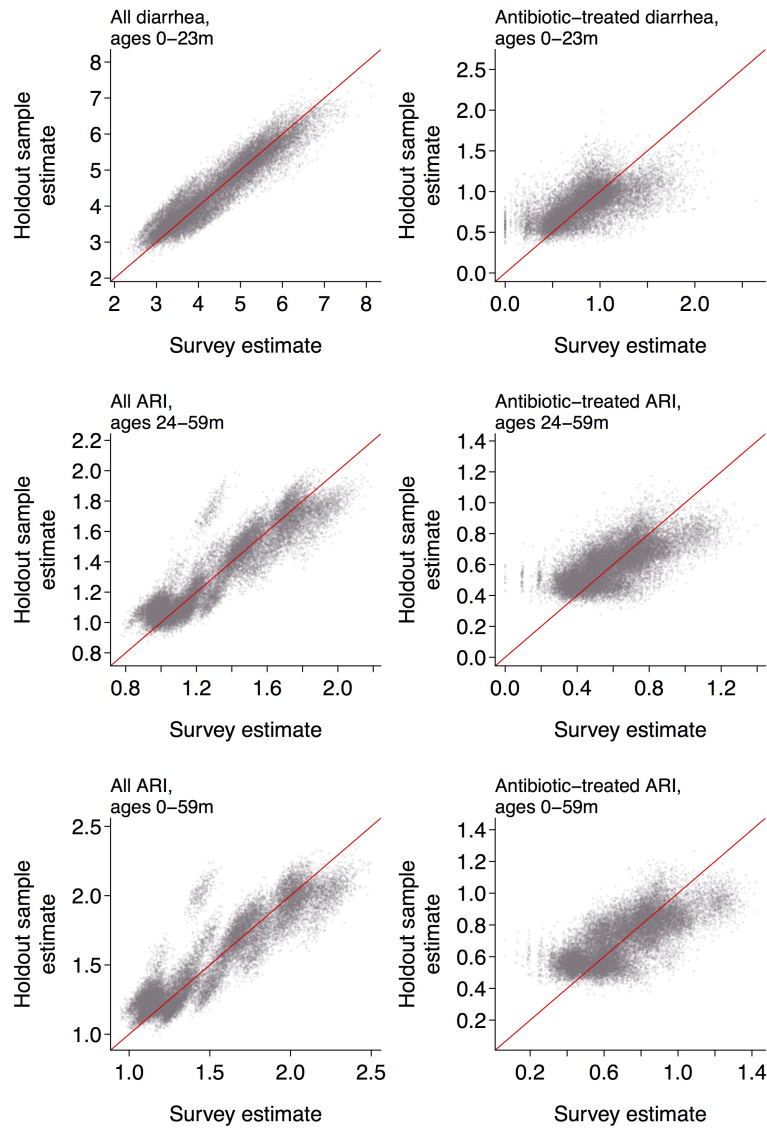
individual and setting-level risk factors; plotted estimates do not account for the joint distribution of other risk factors with respect to region and GDP per capita. Model parameters are included in Supplementary Table 31; estimates are nearly identical for children aged 24–59 months and 0–59 months (Supplementary Table 32). Estimates are obtained from analyses of DHS and MICS surveys comprising 566,508 children across 77 countries. Quantiles are obtained through 5,000 independent draws from the distribution of estimates.



**Extended Data Fig. 5 | Fitted association of GDP per capita with antibiotic treatment of diarrhoea, by region. a, b,** We illustrate model-estimated probabilities of antibiotic treatment for cases of diarrhoea among children aged 0–23 months in rural (**a**) and urban (**b**) settings associated with region and GDP per capita. Estimates for central and eastern Europe are omitted from plots as wide uncertainty intervals would obscure the illustration of associations in other settings. Shaded regions indicate 95% confidence intervals around estimates. True intercepts are dependent on the distribution

of other individual and setting-level risk factors; plotted estimates do not account for the joint distribution of other risk factors with respect to region and GDP per capita. Model parameter estimates are provided in Supplementary Table 33. Estimates are obtained from analyses of DHS and MICS surveys comprising 377,665 children across 77 countries. Quantiles are obtained through 5,000 independent draws from the distribution of estimates.





**Extended Data Fig. 6 | Out-of-sample performance of the extrapolated incidence estimation.** We illustrate the extrapolated estimates of ARI and diarrhoea incidence and antibiotic-treated ARI and antibiotic-treated diarrhoea incidence against model-based estimates within the 10% holdout

samples reserved to assess out-of-sample performance of predictive modelling. Red lines indicate expectations under the scenario of one-to-one correspondence between true and predicted out-of-sample values.

## Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

### Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

- |                                     |  |
|-------------------------------------|--|
| n/a                                 | Confirmed  |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement  |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly  |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> The statistical test(s) used AND whether they are one- or two-sided<br><i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i>   |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> A description of all covariates tested   |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons  |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> For null hypothesis testing, the test statistic (e.g. $F$ , $t$ , $r$ ) with confidence intervals, effect sizes, degrees of freedom and $P$ value noted<br><i>Give <math>P</math> values as exact values whenever suitable.</i>                            |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings  |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes   |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> Estimates of effect sizes (e.g. Cohen's $d$ , Pearson's $r$ ), indicating how they were calculated   |

*Our web collection on [statistics for biologists](#) contains articles on many of the points above.*

### Software and code

Policy information about [availability of computer code](#)

#### Data collection

Our analysis used publicly-available data sources from the MICS and DHS surveys. MICS data are publicly available at <http://mics.unicef.org/>. DHS data are available upon request at <https://dhsprogram.com/Data/>.

#### Data analysis

Analyses were conducted in R (version 3.5.1). We used the "metafor" package (for meta-analysis); the "survival" package (for fitting of conditional logistic regression models); the "Amelia II" package (for multiple imputation); and the "caret" and "gbm" packages (for cross-validation analyses). Analysis code is available at <https://github.com/joelewnard/global-abx/>

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

### Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Analysis code is available from <http://github.com/joelewnard> to replicate all parts of the analysis and generation of figures. MICS data are publicly available at <http://mics.unicef.org/>. DHS data are available upon request at <https://dhsprogram.com/Data/>. Files posted to the author's github repository above include reduced versions of the DHS and MICS datasets necessary to replicate analyses.

## Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☒ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://nature.com/documents/nr-reporting-summary-flat.pdf)

## Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	We conducted a case-control study of the association of pneumococcal conjugate vaccine and rotavirus vaccine receipt with mother-reported acute respiratory infection and diarrhea and related care-seeking among children under five years of age in low- and middle-income countries, and estimated the vaccine-preventable burden of antibiotic use associated with these infections.
Research sample	Children under five years of age residing in low- and middle-income countries, whose mothers responded to Demographic Health Survey (DHS) and Multiple Indicator Cluster Survey (MICS) rounds. Children under five years old are of interest as the population at greatest risk of antibiotic-treated ARI and diarrhea episodes which may be preventable by pneumococcal conjugate vaccines and rotavirus vaccines. The DHS and MICS surveys address household composition and risk factors, health outcomes, and healthcare utilization among household occupants. The design of the DHS and MICS sampling strategies provides a nationally representative selection of households, enabling valid estimation of incidence rates and probabilities of care seeking or antibiotic receipt, given illness; data were available from 77 countries, in total, for our analyses. The DHS/MICS studies provide the largest sample of individual-level observations on ARI and diarrhea outcomes, vaccination status, and risk factors among children in LMICs.
Sampling strategy	DHS and MICS use a probability-based geographic sampling scheme to select clusters within countries (or provinces of countries), and households within clusters; as our analyses were of a secondary nature, we could not dictate statistical power by direct enrollment of study subjects, and used data available from all children. For the case-control study, data were available on 65,815 children, including 5342 ARI cases and 9944 diarrhea cases. Based on the number of children available for analysis, we expected sufficient statistical power to estimate vaccine direct effects against all-cause disease endpoints under a 1:3 matched case-control design. Our analysis of risk factor data to estimate incidence rates included all 944,173 children for whom data were available from DHS/MICS surveys beginning in 2006; in this large sample, we likewise expected sufficient statistical power would be available for estimation of the effects of individual risk factors on ARI and diarrhea outcomes.
Data collection	We conducted secondary analyses of the DHS and MICS datasets; data collection for the original surveys is conducted using standardized, pre-piloted questionnaires and is described elsewhere (see <a href="https://dhsprogram.com/">https://dhsprogram.com/</a> and <a href="http://mics.unicef.org/">http://mics.unicef.org/</a> ).
Timing	We used data collected in surveys administered from January 1, 2006 to December 31, 2018. For the case-control study, we used data collected from January 1, 2015 onward; prior surveys did not collect card-confirmed rotavirus and/or pneumococcal vaccination status.
Data exclusions	For the case-control study, we excluded children from whom the following data were unavailable: outcome (ARI or diarrhea and related care-seeking), exposure (vaccination status), and covariates used in the matching procedure (including age, visit timing, country, urbanicity, household wealth quintile, and pentavalent vaccine doses received). For other analyses (development of the burden model), we limited data to the most recent DHS/MICS survey round undertaken in each country and did not exclude children with missing data; missing variables were multiply imputed using the Amelia II package in R.
Non-participation	Our analyses included data from all mothers who consented to participate in DHS/MICS survey rounds; participation in the original surveys was voluntary, and respondents were free to decline to participate or to refuse to answer any question. We conducted multiple imputation to analyze answers to questions for which answers were not available (because the respondent refused to answer or was unsure of the answer).
Randomization	Our study was not randomized. For the case-control analyses, we controlled for the following variables via exact matching: age, visit timing, country, urbanicity, household wealth quintile, and pentavalent vaccine doses received. In addition, we validated estimates by testing for vaccine effects against negative control conditions using the same analysis framework (PCV effects against diarrhea outcomes, and rotavirus vaccine effects against ARI outcomes). Our burden model aimed to estimate incidence rates as a function of multiple individual risk factors, for which causal inference was not the focus of analyses.

## Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

- |                                     |  |
|-------------------------------------|--|
| n/a                                 | Involvement in the study                             |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Antibodies                  |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Eukaryotic cell lines       |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Palaeontology               |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Animals and other organisms |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Human research participants |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Clinical data               |

Methods

- |                                     |   |
|-------------------------------------|---|
| n/a                                 | Involvement in the study                        |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> ChIP-seq               |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Flow cytometry         |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> MRI-based neuroimaging |



# Autophagy promotes immune evasion of pancreatic cancer by degrading MHC-I

<https://doi.org/10.1038/s41586-020-2229-5>

Received: 27 August 2019

Accepted: 28 March 2020

Published online: 22 April 2020

 Check for updates

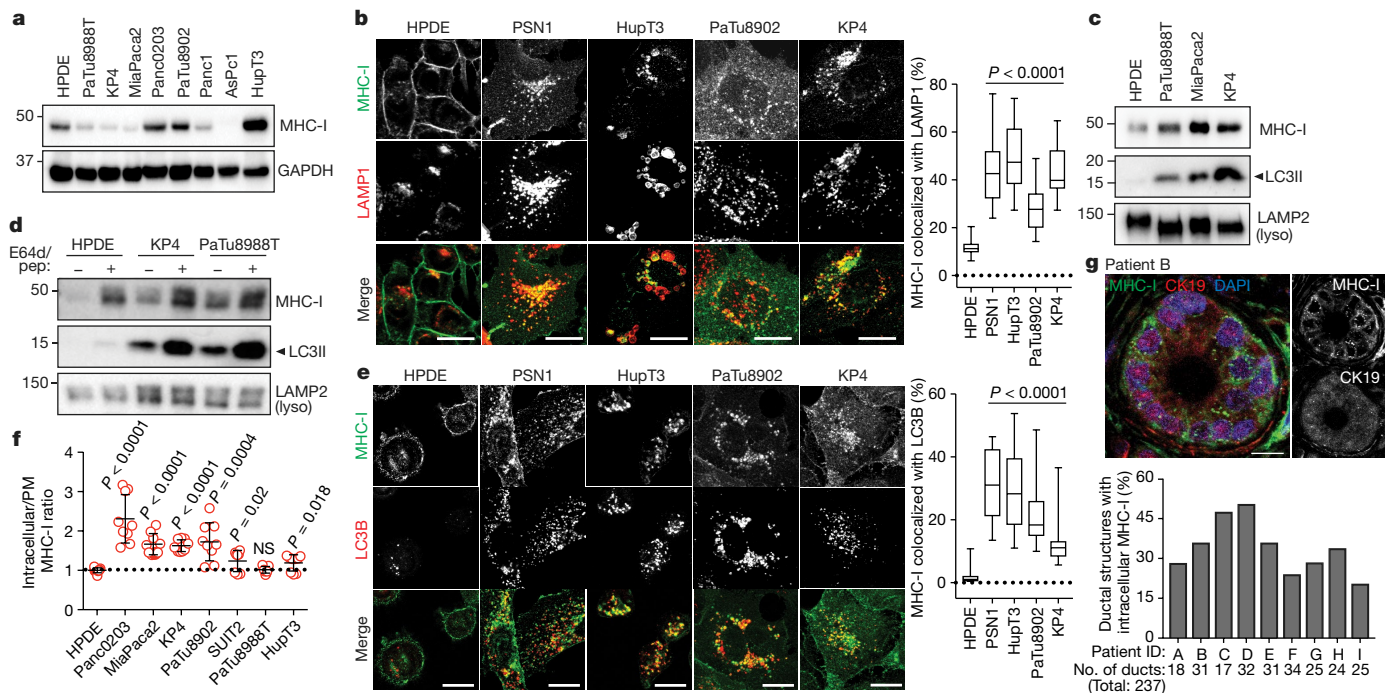
Keisuke Yamamoto<sup>1,12</sup>, Anthony Venida<sup>2,12</sup>, Julian Yano<sup>2</sup>, Douglas E. Biancur<sup>1</sup>, Miwako Kakiuchi<sup>3,4</sup>, Suprit Gupta<sup>2</sup>, Albert S. W. Sohn<sup>1</sup>, Subhadip Mukhopadhyay<sup>1</sup>, Elaine Y. Lin<sup>1</sup>, Seth J. Parker<sup>1</sup>, Robert S. Banh<sup>1</sup>, Joao A. Paulo<sup>5</sup>, Kwun Wah Wen<sup>6</sup>, Jayanta Debnath<sup>6,7</sup>, Grace E. Kim<sup>6</sup>, Joseph D. Mancias<sup>8</sup>, Douglas T. Fearon<sup>9,10,11</sup>, Rushika M. Perera<sup>2,6,7</sup>✉ & Alec C. Kimmelman<sup>1</sup>✉

Immune evasion is a major obstacle for cancer treatment. Common mechanisms of evasion include impaired antigen presentation caused by mutations or loss of heterozygosity of the major histocompatibility complex class I (MHC-I), which has been implicated in resistance to immune checkpoint blockade (ICB) therapy<sup>1–3</sup>. However, in pancreatic ductal adenocarcinoma (PDAC), which is resistant to most therapies including ICB<sup>4</sup>, mutations that cause loss of MHC-I are rarely found<sup>5</sup> despite the frequent downregulation of MHC-I expression<sup>6–8</sup>. Here we show that, in PDAC, MHC-I molecules are selectively targeted for lysosomal degradation by an autophagy-dependent mechanism that involves the autophagy cargo receptor NBRI. PDAC cells display reduced expression of MHC-I at the cell surface and instead demonstrate predominant localization within autophagosomes and lysosomes. Notably, inhibition of autophagy restores surface levels of MHC-I and leads to improved antigen presentation, enhanced anti-tumour T cell responses and reduced tumour growth in syngeneic host mice. Accordingly, the anti-tumour effects of autophagy inhibition are reversed by depleting CD8<sup>+</sup> T cells or reducing surface expression of MHC-I. Inhibition of autophagy, either genetically or pharmacologically with chloroquine, synergizes with dual ICB therapy (anti-PD1 and anti-CTLA4 antibodies), and leads to an enhanced anti-tumour immune response. Our findings demonstrate a role for enhanced autophagy or lysosome function in immune evasion by selective targeting of MHC-I molecules for degradation, and provide a rationale for the combination of autophagy inhibition and dual ICB therapy as a therapeutic strategy against PDAC.

Human PDAC cell lines expressed heterogeneous levels of total MHC-I protein (Fig. 1a), and notably, exhibited a punctate cytoplasmic distribution of MHC-I that co-localized with lysosomes (Fig. 1b). By contrast, non-transformed human pancreatic ductal epithelial (HPDE) cells showed predominant localization of MHC-I on the plasma membrane (Fig. 1b). Indeed, MHC-I molecules were highly enriched in PDAC lysosomes as compared to HPDE lysosomes (Fig. 1c, Extended Data Fig. 1a, b). Moreover, lysosomal inhibition resulted in accumulation of MHC-I within lysosomes, which confirms that MHC-I is actively routed to the lysosome for degradation (Fig. 1d). A substantial fraction of the MHC-I puncta also co-localized with LC3B-labelled autophagosomes in PDAC cells, consistent with the increased autophagy levels in PDAC<sup>9–11</sup> (Fig. 1e). Notably, similar phenotypes were observed in

several non-small-cell lung cancer (NSCLC) cell lines (Extended Data Fig. 1c, d). Flow cytometry-based analysis of total intracellular versus plasma membrane MHC-I confirmed a higher relative abundance of intracellular MHC-I in most PDAC cell lines (Fig. 1f). Similarly, surface levels of MHC-I were lower in PDAC cells derived from a genetically engineered mouse model (GEMM) of PDAC<sup>12</sup> than in normal pancreas cells (Extended Data Fig. 1e). Furthermore, immunofluorescence staining revealed that all human PDAC tumours analysed contained considerable regions with intracellular MHC-I localization (Fig. 1g, Extended Data Fig. 1f), supporting our *in vitro* findings. Together, these data suggest that MHC-I molecules are reduced at the cell surface and predominantly localized within autophagosomes and lysosomes in PDAC. Autophagy inhibition by ATG3 and ATG7 knockdown as well as

<sup>1</sup>Department of Radiation Oncology, Perlmutter Cancer Center, New York University School of Medicine, New York, NY, USA. <sup>2</sup>Department of Anatomy, University of California, San Francisco, San Francisco, CA, USA. <sup>3</sup>Columbia Center for Translational Immunology, Columbia University College of Physicians and Surgeons, New York, NY, USA. <sup>4</sup>Columbia Stem Cell Initiative, Columbia University College of Physicians and Surgeons, New York, NY, USA. <sup>5</sup>Department of Cell Biology, Harvard Medical School, Boston, MA, USA. <sup>6</sup>Department of Pathology, University of California, San Francisco, San Francisco, CA, USA. <sup>7</sup>Helen Diller Family Comprehensive Cancer Center, University of California, San Francisco, San Francisco, CA, USA. <sup>8</sup>Division of Radiation and Genome Stability, Department of Radiation Oncology, Dana-Farber Cancer Institute, Harvard Medical School, Boston, MA, USA. <sup>9</sup>Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, USA. <sup>10</sup>Weill Cornell Medicine, New York, NY, USA. <sup>11</sup>Cancer Research UK Cambridge Institute, University of Cambridge, Robinson Way, Cambridge, UK. <sup>12</sup>These authors contributed equally: Keisuke Yamamoto, Anthony Venida. ✉e-mail: [rushika.perera@ucsf.edu](mailto:rushika.perera@ucsf.edu); [alec.kimmelman@nyulangone.org](mailto:alec.kimmelman@nyulangone.org)



**Fig. 1 | MHC-I is enriched in lysosomes of PDAC cells and displays reduced cell surface expression.** **a**, Levels of MHC-I (HLA-A, -B, -C) in HPDE and human PDAC cell lines. **b**, Left, localization of MHC-I (green) relative to LAMP1-positive (red) lysosomes. Scale bars, 20  $\mu$ m. Right, the percentage co-localization ( $n = 14$ –20 fields). **c**, Presence of MHC-I in immuno-isolated lysosomes. **d**, Accumulation of MHC-I in immuno-isolated lysosomes after treatment with lysosome inhibitors E64d and pepstatin A (pep) for 6 h. **e**, Left, localization of MHC-I (green) relative to LC3B-positive (red) autophagosomes. Scale bars, 20  $\mu$ m. Right, the percentage co-localization ( $n = 14$ –20 fields). **f**, Flow cytometry-based analysis of intracellular versus plasma membrane (PM) MHC-I

levels. Graph shows higher intracellular MHC-I relative to plasma membrane MHC-I in PDAC cells ( $n = 9$  replicates pooled from 3 independent experiments per cell line). Data are mean  $\pm$  s.d. **g**, Top, intracellular localization of MHC-I (green) in CK19-positive (red) ducts from a specimen from a patient with PDAC. Scale bars, 20  $\mu$ m. Bottom, the percentage of ducts showing intracellular MHC-I localization. A representative of at least two independent experiments is shown in **a**, **c**, **d**. For box-and-whisker plots (**b**, **e**), centre lines indicate median values and whiskers represent minimum and maximum values.  $P$  values were determined by unpaired two-tailed  $t$ -tests. NS, not significant. See Source Data for exact  $n$  values for **b**, **e**. See Supplementary Fig. 1 for gel source data.

lysosomal inhibition with V-ATPase inhibitor bafilomycin A1 (BafA1) increased total and plasma membrane MHC-I levels in PDAC cells (Fig. 2a, b, Extended Data Fig. 2a–i). Moreover, surface MHC-I levels of *Atg5*<sup>−/−</sup> mouse PDAC cells<sup>10</sup> were higher than those of *Atg5*<sup>+/+</sup> PDAC cells (Extended Data Fig. 2j). Notably, lysosomal inhibition with BafA1 or chloroquine increased levels of MHC-I proteins but did not affect those involved in antigen processing and presentation (Extended Data Fig. 2k, l), which suggests that autophagy inhibition would not impair these steps. Similar phenotypes were also observed in several NSCLC cell lines (Extended Data Fig. 2m–o).

Previous studies have revealed non-canonical functions for a subset of autophagy proteins in alternative trafficking pathways such as LC3-associated phagocytosis (LAP)<sup>13</sup> and endocytosis (LANDO)<sup>14</sup>. Knockdown of FIP200, which is required for autophagy but not for LAP or LANDO, led to an increase in MHC-I levels (Extended Data Fig. 3a–c). Similarly, knockdown of ATG14, ATG13 or ULK1, all required for autophagy but dispensable for LAP, increased surface MHC-I levels (Extended Data Fig. 3d–j). By contrast, no obvious increase in surface MHC-I was observed after knockdown of RUBICON, which is required for LAP and LANDO but not for autophagy (Extended Data Fig. 3k, l). Collectively, these data suggest a specific role for macroautophagy in the trafficking of MHC-I to the lysosome.

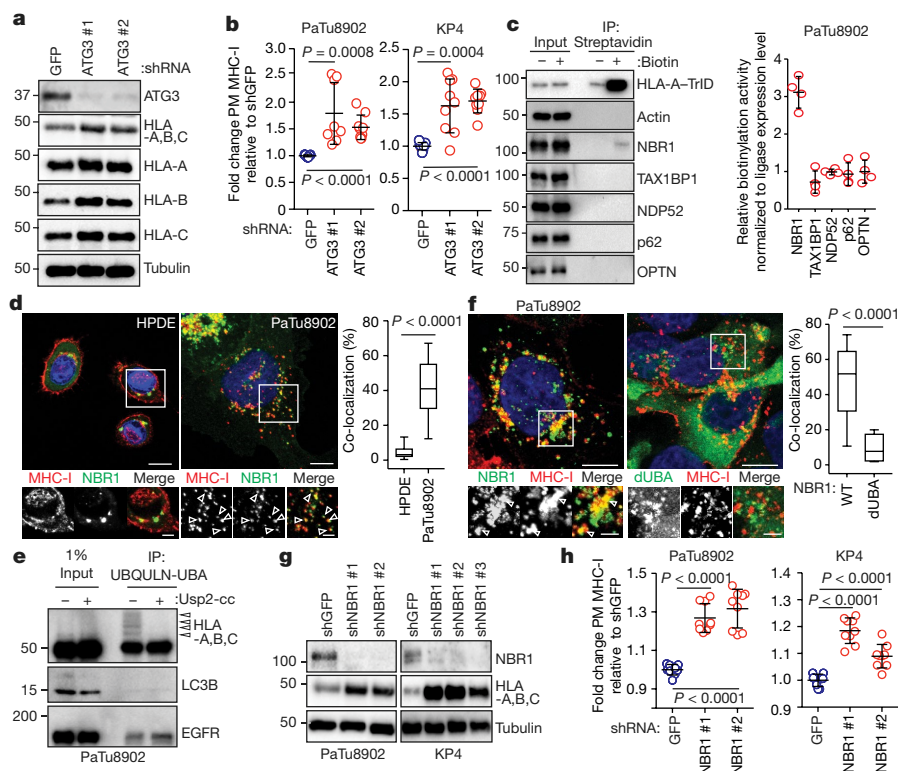
### NBR1 mediates selective autophagy of MHC-I

Autophagy can selectively degrade target molecules using autophagy cargo receptor proteins that bind to and recruit substrates to autophagosomal membranes<sup>15</sup>. To identify autophagy receptor protein(s) involved in MHC-I degradation in PDAC cells, we generated

a proximity biotinylation assay in which the C terminus of HLA-A was fused to the biotin ligase TurboID<sup>16</sup> and Flag (HLA-A–TrID). After addition of biotin, HLA-A–TrID covalently tags endogenous proteins within a few nanometres of the ligase with biotin (Extended Data Fig. 4a, b). Among the autophagy receptors tested, only NBR1 showed significant biotinylation, indicating that NBR1 interacts with MHC-I (Fig. 2c). Furthermore, immunofluorescence revealed more frequent co-localization between NBR1 and MHC-I in PDAC cells relative to HPDE cells (Fig. 2d). NBR1 has been shown to interact with and target ubiquitylated substrates for degradation<sup>17</sup>. Indeed, MHC-I is poly-ubiquitylated in PDAC cells (Fig. 2e, Extended Data Fig. 4c) whereas LC3B is not ubiquitylated and EGFR is mono-ubiquitylated, as previously described<sup>18</sup>. Accordingly, NBR1 lacking its ubiquitin-associated (UBA) domain was unable to co-localize with MHC-I, despite retaining localization with LC3B (Fig. 2f, Extended Data Fig. 4d). Finally, NBR1 knockdown increased total and plasma membrane MHC-I levels in PDAC cells (Fig. 2g, h, Extended Data Fig. 4e, f), confirming a role for NBR1 in MHC-I regulation. Together, these data demonstrate that surface MHC-I is decreased in PDAC via an NBR1-mediated autophagy–lysosomal pathway.

### Autophagy inhibition enhances anti-tumour immunity

CD8<sup>+</sup> cytotoxic T cells have crucial roles in anti-tumour immunity. We hypothesized that reduced surface levels of MHC-I on PDAC cells may facilitate their evasion from CD8<sup>+</sup> T cells, which recognize tumour antigens presented by MHC-I. To test this, we used mouse PDAC cells derived from C57BL/6 mice and engineered them to express a doxycycline (Dox)-inducible dominant-negative mutant of ATG4B



**Fig. 2 | NBR1 promotes MHC-I trafficking to the lysosome through an autophagy-dependent pathway.**

**a**, Effect of short hairpin RNA (shRNA)-mediated ATG3 knockdown on levels of HLA-A, HLA-B and HLA-C in human PaTu8902 cells. **b**, Flow cytometry-based quantification of plasma membrane levels of MHC-I (HLA-A, -B, -C) (PaTu8902,  $n = 9$ , 8, 9; KP4,  $n = 9$  per group; data pooled from three independent experiments) after shRNA-mediated ATG3 knockdown. **c**, Left, KP4 cells expressing HLA-A–TrID were labelled with 10  $\mu$ M biotin for 30 min. Biotinylation of proteins was detected after streptavidin pull down. Asterisk denotes self-biotinylation of HLA-A–TrID. IP, immunoprecipitation. Right, enrichment for each receptor ( $n = 4$  independent experiments). **d**, Localization of MHC-I (red) relative to GFP–NBR1 (green) in HPDE and PaTu8902 cells. Arrowheads show examples of co-localization. Scale bars, 20  $\mu$ m. Graph shows quantification of co-localization (HPDE,  $n = 23$  fields; PaTu8902,  $n = 20$  fields). **e**, Endogenous ubiquitinated proteins were affinity captured from PaTu8902 cells with

UBQLN1-UBA-conjugated beads. Arrowheads indicate MHC-I polyubiquitylation. Treatment of affinity-captured samples for 1 h with purified deubiquitylating enzyme Usp2-cc (+) to induce deubiquitylation leads to loss of MHC-I polyubiquitylation. Control proteins: LC3B (no ubiquitylation) and EGFR (mono-ubiquitylation). **f**, Left, endogenous MHC-I co-localizes with wild-type (WT) GFP–NBR1 but not GFP–NBR1 lacking its UBA domain (dUBA). Right, quantification of co-localization (GFP–NBR1;  $n = 17$  fields; GFP–NBR1 dUBA;  $n = 14$  fields). Scale bars, 20  $\mu$ m and 10  $\mu$ m (inset). **g**, Effect of shRNA-mediated NBR1 knockdown on levels of MHC-I. **h**, Flow cytometry-based quantification of plasma membrane MHC-I ( $n = 9$  replicates from three independent experiments) after NBR1 knockdown. A representative of at least two independent experiments is shown in **a**, **e**, **g**. Data are mean  $\pm$  s.d. (**b**, **c**, **h**). Box-and-whisker plots are as in Fig. 1.  $P$  values were determined by unpaired two-tailed  $t$ -tests. See Supplementary Fig. 1 for gel source data.

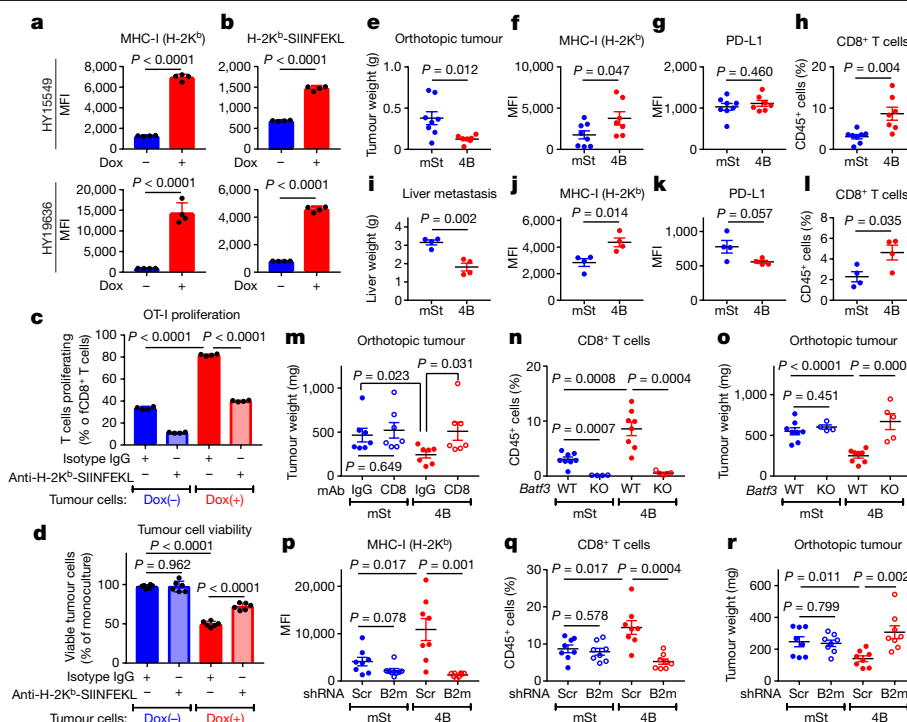
(ATG4B(C74A)), which potently inhibits autophagy<sup>12,19</sup>. Dox treatment efficiently inhibited autophagy and increased surface MHC-I levels (Extended Data Fig. 5a, b) but did not affect other cell-surface molecules, such as those that are immune inhibitory (Extended Data Fig. 5c). These data support a specific mechanism that links surface MHC-I levels and autophagy.

To evaluate antigen presentation, the model neoantigen ovalbumin (OVA) was expressed in mouse PDAC cells carrying Dox-inducible ATG4B(C74A) (Extended Data Fig. 5d). Dox treatment increased surface expression of both MHC-I and OVA-derived peptide SIINFEKL bound to H-2K<sup>b</sup>, confirming enhanced peptide presentation (Fig. 3a, b, Extended Data Fig. 5e, f). OVA-specific CD8<sup>+</sup> T cells (OT-I cells) co-cultured with PDAC cells pre-treated with Dox showed higher proliferation and expression of IFN- $\gamma$  and TNF than those co-cultured with non-pre-treated PDAC cells (Fig. 3c, Extended Data Fig. 5g, h). Accordingly, Dox-treated PDAC cells showed reduced viability after co-culture with OT-I cells (Fig. 3d), indicating enhanced T-cell-mediated tumour cell killing. Notably, these effects were MHC-I specific, as a H-2K<sup>b</sup>-SIINFEKL blocking antibody partially inhibited OT-I proliferation and rescued PDAC cell viability in the setting of autophagy inhibition (Fig. 3c, d). Collectively, these

results indicate that tumour-specific inhibition of autophagy leads to increased antigen presentation, which enhances CD8<sup>+</sup> T cell proliferation, activation and tumour cell killing in vitro.

To test the effect of autophagy inhibition on anti-tumour immune responses in vivo, mouse PDAC cells expressing Dox-inducible mStrawberry (mSt) or mSt-ATG4B(C74A) (4B) (Extended Data Fig. 6a) were orthotopically transplanted into syngeneic (C57BL/6) mice. Autophagy-inhibited cells (4B) formed smaller tumours with higher MHC-I expression than control cells (mSt) while PD-L1 expression was unchanged (Fig. 3e–g, Extended Data Figs. 6b–f, 7k). Moreover, autophagy-inhibited tumours (4B) exhibited a significant increase in infiltrating CD8<sup>+</sup> T cells and a decrease in myeloid-derived suppressor cells, a major immunosuppressive cell type in PDAC tumours<sup>20</sup> (Fig. 3h, Extended Data Fig. 6g–j). Similarly, knockdown of ATG7 resulted in a significant reduction in tumour burden and an increase in tumour-infiltrating T cells (Extended Data Fig. 6k–n). Notably, there was a significant correlation between smaller tumour sizes and increased CD8<sup>+</sup> T cell infiltration (Extended Data Fig. 6o), supporting a role of T cell immunity in control of autophagy-deficient tumours. Similar results were obtained in a liver metastasis model: mice





**Fig. 3 | Autophagy inhibition enhances anti-tumour T cell response.** **a, b**, Surface H-2K<sup>b</sup> (**a**) and H-2K<sup>b</sup>.SIINFEKL (**b**) measured by flow cytometry. Mouse PDAC cells expressing OVA and Dox-inducible mTurquoise2–ATG4B(C74A) were grown as organoids and treated with or without Dox (1 µg ml<sup>-1</sup>) for 96 h (*n* = 4 per group). MFI, mean fluorescent intensity. **c, d**, Co-culture of OT-I cells with HY19636 cells shown in **a** and **b**. After 48 h, OT-I proliferation was measured by CFSE dilution (*n* = 4 per group) (**c**) and PDAC cell viability was measured by Cell-Titer Glo (*n* = 6 per group) (**d**). **e–r**, HY15549 cells carrying Dox-inducible mStrawberry (mSt) or mSt-ATG4B(C74A) (4B) were orthotopically (**e–h, m–r**) or intrasplenically (**i–l**) injected into syngeneic mice (C57BL/6). The expression of MHC-I and PD-L1 on PDAC cells (**f, g, j, k, p**) and tumour-infiltrating CD8<sup>+</sup> T cells (**h, l, n, q**) quantified by flow cytometry.

**e–h**, Orthotopic tumours obtained on day 20 (mSt, *n* = 8; 4B, *n* = 7). **i–l**, Livers obtained on day 15 (*n* = 4 per group). **e, i**, Weight of tumours (**e**) and livers (**i**). **f–h, j–l**, Flow cytometry analysis. **m**, Weight of tumours after treatment with isotype control IgG or neutralizing monoclonal antibody against CD8 (*n* = 7 per group). **n, o**, Tumours in wild-type or *Batf3*<sup>-/-</sup> knockout (KO) mice (*n* = 8, 4 and 5; left to right). **n**, Quantification of CD8<sup>+</sup> T cells. **o**, Tumour weight. **p–r**, Tumours expressing control shRNA (Scr) or shRNA against B2m (*n* = 8 per group) collected on day 20. **p, q**, Flow cytometry analysis. **r**, Tumour weight. Data are mean ± s.d. (**a–d**) or s.e.m. (**e–r**). **n** indicates biological replicates (**a–d**) or individual mice (**e–r**). For **a–m, p–r**, experiments were performed at least twice and representative data of one experiment are shown. *P* values were determined by unpaired two-tailed *t*-tests.

injected with autophagy-inhibited cells (4B) exhibited lower metastatic burden, higher MHC-I expression on cancer cells, and more tumour-infiltrating CD8<sup>+</sup> T cells than mice injected with control cells (mSt) (Fig. 3i–l, Extended Data Fig. 6p–t). Together, these results confirm that autophagy-inhibition restores surface levels of MHC-I on cancer cells and enhances anti-tumour T cell response in vivo.

Notably, antibody-mediated CD8<sup>+</sup> T cell depletion restored the growth of autophagy-inhibited tumours (4B) (Fig. 3m, Extended Data Fig. 7a–c), confirming the role of CD8<sup>+</sup> T cells in tumour control. We also find that mice deficient for CD103<sup>+</sup> dendritic cells (*Batf3*<sup>-/-</sup>)<sup>21</sup>, which have pivotal roles in CD8<sup>+</sup> T cell priming<sup>22</sup> and recruitment into tumours<sup>20,23</sup>, showed almost complete loss of tumour-infiltrating CD8<sup>+</sup> T cells (Fig. 3n, Extended Data Fig. 7d–f). Accordingly, growth of autophagy-inhibited tumours (4B) was restored in *Batf3*<sup>-/-</sup> mice (Fig. 3o). These results indicate that the anti-tumour effects of tumour-specific autophagy inhibition are mediated, at least in part, by CD8<sup>+</sup> T cells, and this process requires CD103<sup>+</sup> dendritic cells.

Finally, to confirm the effect of increased MHC-I expression after autophagy inhibition on tumour growth in vivo, cell surface levels of MHC-I were depleted by knockdown of beta-2 microglobulin (B2M), a crucial component of the MHC-I complex (Extended Data Fig. 7g). B2M knockdown led to MHC-I depletion in vivo, decreased the number of CD8<sup>+</sup> T cells in autophagy-inhibited tumours (4B), and rescued the growth of autophagy-inhibited tumours (4B) (Fig. 3p–r, Extended Data Fig. 7h–k). We also found that tumour-infiltrating CD103<sup>+</sup> DCs, which are increased in autophagy-inhibited tumours (4B), were decreased

after B2M knockdown (Extended Data Fig. 7f, j). This is in line with a recent study showing that MHC-I-restricted antigen-recognition by CD8<sup>+</sup> T cells can trigger activation of dendritic cells, which further augments recruitment of CD8<sup>+</sup> T cells into the tumour<sup>24</sup>. Overall, these data indicate that increased surface expression of MHC-I on PDAC cells after tumour-specific inhibition of autophagy is a prerequisite for increased CD8<sup>+</sup> T cell infiltration and tumour cell killing.

## Basal autophagy determines immunogenicity

Our data led us to hypothesize that basal autophagy flux might determine the immunogenicity of PDAC cells. Mouse PDAC cells grown as organoids<sup>25</sup> exhibited considerable heterogeneity in autophagy flux (measured as the ratio of green to red fluorescent protein signal, GFP/RFP), which closely resembled that of orthotopic tumours (Extended Data Fig. 8a, b). Taking advantage of this heterogeneity, PDAC cells with the lowest and highest 20% of GFP/RFP ratio were isolated as the autophagy-high (ATHi) and autophagy-low (ATLo) cells (Extended Data Fig. 8c). Transcriptome analysis confirmed upregulation of autophagy or lysosome related genes and the Mit/TFE transcription factors, master regulators of autophagy or lysosome gene expression<sup>11</sup>, in the ATHi population (Extended Data Fig. 8d, e). Notably, ATHi and ATLo cells derived from a clonal population showed a similar transcriptional profile (Extended Data Fig. 8f), suggesting that diversities in basal autophagy flux arise not from genetic variations but instead from heterogeneous expression of the autophagy or lysosome gene program<sup>11</sup>.

Consistent with the cell-autonomous roles of autophagy in PDAC<sup>9,10</sup>, ATlo cells exhibited reduced clonogenic capacity in vitro (Extended Data Fig. 8g, h), which is a similar result to the decreased clonogenic growth of ATG4B(C74A)-expressing cells (Extended Data Fig. 8i). After orthotopic transplantation into syngeneic mouse hosts, ATlo cells gave rise to smaller tumours than ATHi cells (Extended Data Fig. 9a). Notably, this was reproduced using cells sorted from a clonal population (Extended Data Fig. 9b), which suggests that tumour-cell intrinsic factors account for the observed phenotypes rather than differences in retroviral-vector integration sites or copy numbers of the reporter. Furthermore, ATlo-derived tumours exhibited higher MHC-I expression and more tumour-infiltrating CD8<sup>+</sup> T cells, which inversely correlated with tumour weight, compared to ATHi-derived tumours (Extended Data Fig. 9c–f). This growth advantage of ATHi-derived tumours was lost in nude mice (Extended Data Fig. 9g), confirming a role of T cells in the suppression of ATlo-derived tumours. In the liver metastasis model, mice injected with ATlo cells exhibited lower metastatic burden and higher MHC-I expression on PDAC cells than mice injected with ATHi cells (Extended Data Fig. 9h–j). CD8<sup>+</sup> T cells tended to accumulate around PDAC cells with low-autophagy flux, whereas CD8<sup>+</sup> T cells were scarce around PDAC cells with high-autophagy flux (Extended Data Fig. 9k). These results, along with the data in autophagy inhibition models, indicate that autophagy is a crucial determinant of immunogenicity in PDAC cells.

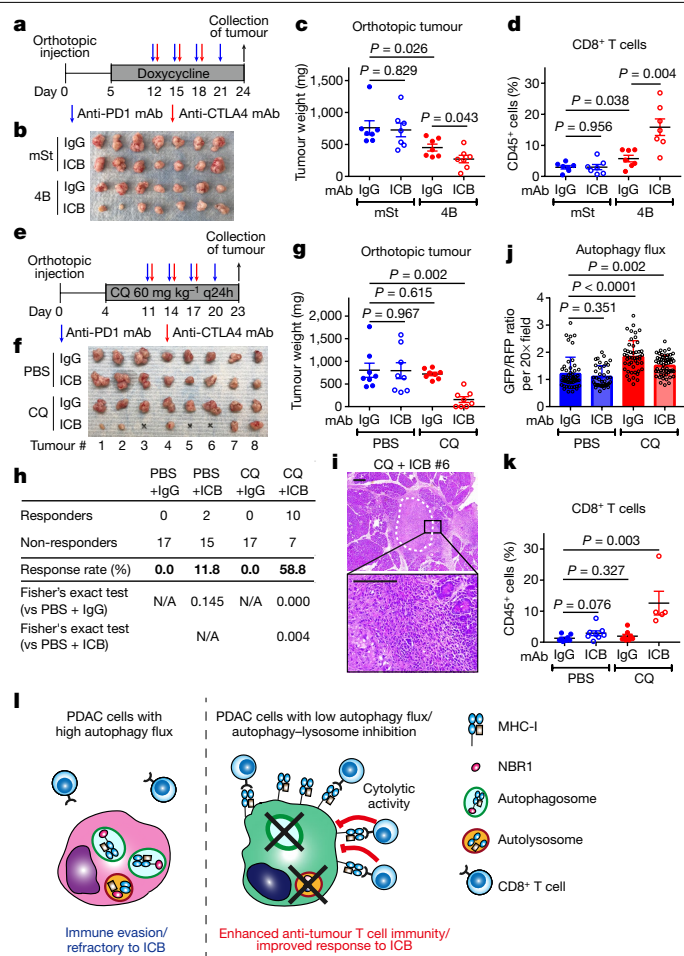
### Autophagy inhibition sensitizes PDAC to dual ICB

PDAC is refractory to ICB<sup>4</sup>. To test whether inhibition of autophagy might sensitize PDAC to ICB, we treated established syngeneic orthotopic tumours with anti-PD-1 antibody alone or dual ICB (anti-PD-1 and anti-CTLA4 antibodies). Consistent with a recent clinical trial<sup>4</sup>, control tumours (mSt) did not respond to either treatment. By contrast, autophagy-inhibited tumours (4B) responded significantly to dual ICB (Fig. 4a–c), but not to anti-PD-1 antibody alone (Extended Data Fig. 10a–c). These tumours showed increased infiltration with bulk CD8<sup>+</sup> T cells and PD1<sup>+</sup>TIM3<sup>+</sup> cells (Fig. 4d, Extended Data Fig. 10d–g), which retain greater function than the most severely exhausted or dysfunctional PD1<sup>+</sup>TIM3<sup>+</sup> cells<sup>26</sup>. These results indicate that tumour-specific autophagy inhibition sensitizes PDAC tumours to dual ICB.

Finally, we assessed the translatability of our findings to systemic autophagy inhibition using chloroquine—a clinically available anti-malarial agent that inhibits acidification of the lysosome and has been used to inhibit autophagy in patients<sup>27–29</sup>. Treatment with lysosomal inhibitors chloroquine or BafA1 increased surface levels of MHC-I in mouse PDAC cells in vitro (Extended Data Fig. 10h). Chloroquine treatment also increased surface expression of MHC-I, but not PD-L1, in orthotopic tumours (Extended Data Fig. 10i–k). However, chloroquine monotherapy failed to significantly reduce tumour weight or increase the infiltration of T cells (Extended Data Fig. 10l, m), possibly due to the unfavourable pharmacokinetics of this drug in vivo<sup>27</sup>. Notably, the combination of chloroquine and dual ICB exerted potent anti-tumour activity (Fig. 4e–i) and a reduction in autophagy flux was confirmed in chloroquine-treated tumours (Fig. 4j, Extended Data Fig. 10n). Moreover, tumours treated with chloroquine plus ICB exhibited increased infiltrations of CD8<sup>+</sup> T cells (Fig. 4k) and an increase in the number of functional PD1<sup>+</sup>TIM3<sup>+</sup>CD8<sup>+</sup> T cells (Extended Data Fig. 10o). Overall, these data indicate that autophagy inhibition, either in cancer cells alone or systemically, sensitizes PDAC tumours to dual ICB therapy (Fig. 4l).

### Discussion

Our results suggest that autophagy is a crucial regulator of immunogenicity in PDAC cells. This is in line with a recent study that showed that the lysosomal pathway is strongly correlated with reduced infiltration



**Fig. 4 | Autophagy inhibition sensitizes PDAC to dual ICB.** **a–d**, Mice bearing orthotopic tumours (HY15549) expressing Dox-inducible mSt or 4B received isotype control IgG or dual ICB (anti-PD1/CTLA4 monoclonal antibodies; mAb) ( $n = 7$  per group). **a**, Study design. **b, c**, Images (**b**) and weight (**c**) of tumours. **d**, Quantification of tumour-infiltrating CD8<sup>+</sup> T cells by flow cytometry. **e–h**, Mice bearing orthotopic tumours (HY15549) expressing the GFP-LC3-RFP reporter received chloroquine (CQ) and ICB ( $n = 8$  per group). **e**, Study design. q24h, every 24 hours. **f, g**, Images (**f**) and weight (**g**) of tumours. No macroscopic tumour was identified in three mice receiving chloroquine plus dual ICB (#3, 5 and 6). **h**, Response rates from two independent experiments. Response is defined as more than 80% reduction in tumour weight as compared with control tumours (PBS + IgG). N/A, not applicable. **i**, Representative haematoxylin and eosin (H&E) images of the pancreas undergoing tumour regression (#6). White dashed line indicates tumour remnants. Scale bars, 250 and 100  $\mu$ m (inset). **j**, Autophagy flux represented by GFP/RFP ratio per 20  $\times$  field ( $n = 49, 39, 47$  and 54; left to right). Increased GFP/RFP ratio indicates reduced autophagy flux. **k**, Quantification of tumour-infiltrating CD8<sup>+</sup> T cells by flow cytometry ( $n = 8, 8, 8$  and 5; left to right). **l**, In PDAC cells, surface MHC-I is downregulated by active degradation through the autophagy-lysosome system, contributing to the primary resistance to ICB. NBR1 binds to MHC-I, facilitating its trafficking to autophagosomes (left). Inhibition of autophagy or the lysosome restores surface MHC-I expression, leading to enhanced anti-tumour T cell immunity and improved response to ICB (right). Data are mean  $\pm$  s.e.m. (**c, d, g, k**) or s.d. (**j**).  $n$  indicates individual mice (**c, d, g, k**) or individual 20 $\times$  fields (**j**). All experiments were performed twice and representative data of one experiment are shown.  $P$  values were determined by unpaired two-tailed  $t$ -tests.

of CD8<sup>+</sup> T cells in human PDAC<sup>30</sup>. In addition, autophagy-related genes are enriched in MHC-I negative PDAC cells that reside in liver metastasis<sup>8</sup>, also suggesting a role for autophagy as a negative regulator of MHC-I. Given the critical roles for the autophagy and lysosome



system in supporting PDAC metabolism and growth<sup>9,11</sup>, our data on immune evasion add to the growing list of cell-autonomous and non-cell-autonomous functions of the autophagy and lysosome system in PDAC pathogenesis<sup>10,12</sup>.

We found that systemic autophagy inhibition by chloroquine, as well as tumour-specific autophagy inhibition, sensitizes PDAC to dual ICB. In addition, recent evidence accounts for the improved therapeutic effects of systemic autophagy inhibition. First, host autophagy supports tumour growth by providing nutrients<sup>31–33</sup>. Second, loss of autophagy proteins or chloroquine treatment increases surface levels of MHC-I in dendritic cells, leading to enhanced CD8<sup>+</sup> T cell response in models of virus infection<sup>34</sup>. Also, autophagy inhibition directly enhances anti-tumour activity of CD8<sup>+</sup> T cells<sup>35</sup>. Finally, loss of LAP, a process also inhibited by chloroquine, polarizes tumour-associated macrophages into a tumour-suppressive phenotype (M1), promoting anti-tumour T cell responses<sup>13</sup>. In line with these, clinical trials have shown that hydroxychloroquine, a derivative of chloroquine, has activity in patients with PDAC<sup>28,29</sup>. Whether the addition of ICB would be synergistic remains to be determined.

In this study, we focused on CD8<sup>+</sup> T cells, given their direct interaction with MHC-I on cancer cells. However, we also observed changes in other immune cells such as myeloid-derived suppressor cells, CD4<sup>+</sup> T cells and CD103<sup>+</sup> dendritic cells after inhibition of autophagy. Investigating how these and changes in other immune cell types are mediated, such as the potential involvement of secreted proteins from PDAC cells after autophagy inhibition, will be important subjects of future work.

Despite the evidence mentioned above and previous work demonstrating that inhibition of autophagy does not impair anti-tumour adaptive immunity<sup>36</sup>, autophagy inhibition has also been reported to effect some aspects of the immune system such as memory formation in virus-specific CD8<sup>+</sup> T cells and CD4<sup>+</sup> T cell priming by dendritic cells<sup>37,38</sup> and may impair chemotherapy-induced immunogenic cell death<sup>39</sup>. Therefore, further studies will be needed to define more subtle aspects of the immune response after autophagy inhibition and how to best combine autophagy or lysosome blockade with cytotoxic and immune-based therapies. For example, deeper insights into the potential effect of basal immunogenicity<sup>40</sup> on response to autophagy inhibition, the differences in mutational burden in mouse and human tumours<sup>41</sup>, the effect of potential dominant antigens in experimental systems<sup>41</sup>, and the heterogeneity described here in terms of MHC-I expression in patients will probably contribute to the successful clinical translation of our findings. Importantly, the mechanistic insights described here about how autophagy can promote immune evasion, provide strong rationale to pursue these studies with the ultimate goal of developing new therapeutic approaches for patients with PDAC.

## Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-020-2229-5>.

- Rooney, M. S., Shukla, S. A., Wu, C. J., Getz, G. & Hacohen, N. Molecular and genetic properties of tumors associated with local immune cytolytic activity. *Cell* **160**, 48–61 (2015).
- McGranahan, N. et al. Allele-specific HLA loss and immune escape in lung cancer evolution. *Cell* **171**, 1259–1271 (2017).
- Rodig, S. J. et al. MHC proteins confer differential sensitivity to CTLA-4 and PD-1 blockade in untreated metastatic melanoma. *Sci. Transl. Med.* **10**, eaar3342 (2018).
- O'Reilly, E. M. et al. Durvalumab with or without tremelimumab for patients with metastatic pancreatic ductal adenocarcinoma: a phase 2 randomized clinical trial. *JAMA Oncol.* **5**, 1431–1438 (2019).

- Waddell, N. et al. Whole genomes redefine the mutational landscape of pancreatic cancer. *Nature* **518**, 495–501 (2015).
- Ryschich, E. et al. Control of T-cell-mediated immune response by HLA class I in human pancreatic carcinoma. *Clin. Cancer Res.* **11**, 498–504 (2005).
- Pandha, H., Rigg, A., John, J. & Lemoine, N. Loss of expression of antigen-presenting molecules in human pancreatic cancer and pancreatic cancer cell lines. *Clin. Exp. Immunol.* **148**, 127–135 (2007).
- Pommier, A. et al. Unresolved endoplasmic reticulum stress engenders immune-resistant, latent pancreatic cancer metastases. *Science* **360**, eaao4908 (2018).
- Yang, S. et al. Pancreatic cancers require autophagy for tumor growth. *Genes Dev.* **25**, 717–729 (2011).
- Yang, A. et al. Autophagy is critical for pancreatic tumor growth and progression in tumors with p53 alterations. *Cancer Discov.* **4**, 905–913 (2014).
- Perera, R. M. et al. Transcriptional control of autophagy-lysosome function drives pancreatic cancer metabolism. *Nature* **524**, 361–365 (2015).
- Yang, A. et al. Autophagy sustains pancreatic cancer growth through both cell-autonomous and nonautonomous mechanisms. *Cancer Discov.* **8**, 276–287 (2018).
- Cunha, L. D. et al. LC3-associated phagocytosis in myeloid cells promotes tumor immune tolerance. *Cell* **175**, 429–441 (2018).
- Heckmann, B. L. et al. LC3-associated endocytosis facilitates beta-amyloid clearance and mitigates neurodegeneration in murine Alzheimer's disease. *Cell* **178**, 536–551 (2019).
- Kirkin, V., McEwan, D. G., Novak, I. & Dikic, I. A role for ubiquitin in selective autophagy. *Mol. Cell* **34**, 259–269 (2009).
- Branon, T. C. et al. Efficient proximity labeling in living cells and organisms with TurboID. *Nat. Biotechnol.* **36**, 880–887 (2018).
- Kirkin, V. et al. A role for NBR1 in autophagosomal degradation of ubiquitinated substrates. *Mol. Cell* **33**, 505–516 (2009).
- Mosesson, Y. et al. Endocytosis of receptor tyrosine kinases is driven by monoubiquitylation, not polyubiquitylation. *J. Biol. Chem.* **278**, 21323–21326 (2003).
- Fujita, N. et al. An Atg4B mutant hampers the lipidation of LC3 paralogs and causes defects in autophagosome closure. *Mol. Biol. Cell* **19**, 4651–4659 (2008).
- Li, J. et al. Tumor cell-intrinsic factors underlie heterogeneity of immune cell infiltration and response to immunotherapy. *Immunity* **49**, 178–193 (2018).
- Hildner, K. et al. Batf3 deficiency reveals a critical role for CD8a<sup>+</sup> dendritic cells in cytotoxic T cell immunity. *Science* **322**, 1097–1100 (2008).
- Broz, M. L. et al. Dissecting the tumor myeloid compartment reveals rare activating antigen-presenting cells critical for T cell immunity. *Cancer Cell* **26**, 638–652 (2014).
- Spranger, S., Dai, D., Horton, B. & Gajewski, T. F. Tumor-residing Batf3 dendritic cells are required for effector T cell trafficking and adoptive T cell therapy. *Cancer Cell* **31**, 711–723 (2017).
- Dangaj, D. et al. Cooperation between constitutive and inducible chemokines enables T cell engraftment and immune attack in solid tumors. *Cancer Cell* **35**, 885–900 (2019).
- Boj, S. F. et al. Organoid models of human and mouse ductal pancreatic cancer. *Cell* **160**, 324–338 (2015).
- Sakuishi, K. et al. Targeting Tim-3 and PD-1 pathways to reverse T cell exhaustion and restore anti-tumor immunity. *J. Exp. Med.* **207**, 2187–2194 (2010).
- Munster, T. et al. Hydroxychloroquine concentration-response relationships in patients with rheumatoid arthritis. *Arthritis Rheum.* **46**, 1460–1469 (2002).
- Karasic, T. B. et al. Effect of gemcitabine and nab-paclitaxel with or without hydroxychloroquine on patients with advanced pancreatic cancer: a phase 2 randomized clinical trial. *JAMA Oncol.* **5**, 993–998 (2019).
- Zeh, H. et al. A randomized phase II preoperative study of autophagy inhibition with high-dose hydroxychloroquine and gemcitabine/nab-paclitaxel in pancreatic cancer patients. *Cancer Res.* <https://doi.org/10.1158/1078-0432.CCR-19-4042> (2020).
- Markosyan, N. et al. Tumor cell-intrinsic EPHA2 suppresses anti-tumor immunity by regulating PTGS2 (COX-2). *J. Clin. Invest.* **130**, 3594–3609 (2019).
- Sousa, C. M. et al. Pancreatic stellate cells support tumour metabolism through autophagic alanine secretion. *Nature* **536**, 479–483 (2016).
- Katheder, N. S. et al. Microenvironmental autophagy promotes tumour growth. *Nature* **541**, 417–420 (2017).
- Poillet-Perez, L. et al. Autophagy maintains tumour growth through circulating arginine. *Nature* **563**, 569–573 (2018).
- Loi, M. et al. Macroautophagy proteins control MHC class I levels on dendritic cells and shape anti-viral CD8<sup>+</sup> T cell responses. *Cell Rep.* **15**, 1076–1087 (2016).
- DeVorkin, L. et al. Autophagy regulation of metabolism is required for CD8<sup>+</sup> T cell anti-tumor immunity. *Cell Rep.* **27**, 502–513 (2019).
- Starobinets, H. et al. Antitumor adaptive immunity remains intact following inhibition of autophagy and antimalarial treatment. *J. Clin. Invest.* **126**, 4417–4429 (2016).
- Xu, X. et al. Autophagy is essential for effector CD8<sup>+</sup> T cell survival and memory formation. *Nat. Immunol.* **15**, 1152–1161 (2014).
- Lee, H. K. et al. In vivo requirement for Atg5 in antigen presentation by dendritic cells. *Immunity* **32**, 227–239 (2010).
- Michaud, M. et al. Autophagy-dependent anticancer immune responses induced by chemotherapeutic agents in mice. *Science* **334**, 1573–1577 (2011).
- Moral, J. A. et al. ILC2s amplify PD-1 blockade by activating tissue-specific cancer immunity. *Nature* **579**, 130–135 (2020).
- Ward, J. P., Gubin, M. M. & Schreiber, R. D. The role of neoantigens in naturally occurring and therapeutically induced immune responses to cancer. *Adv. Immunol.* **130**, 25–74 (2016).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2020

# Article

## Methods

### Cell culture

The cell lines PaTu-8988T, KP4, MiaPaca2, Panc 2.03, PaTu-8902, Panc1, AsPc1, HupT3 and A549 were obtained from the American Type Culture Collection (ATCC) or the DSMZ. H441, H358, HCT116 and BEAS-2B were provided by E. Collisson. HPDE was provided by M. Tsao<sup>42</sup>. Cells were cultured in the following media: PaTu-8988T, KP4, MiaPaca2, PaTu-8902, Panc1, AsPc1 and BEAS-2B in DMEM supplemented with 10% FBS; Panc 2.03 and HupT3 in RPMI with 10% FBS; HPDE cells were cultured in keratinocyte serum-free (KSF) medium supplemented by epidermal growth factor and bovine pituitary extract (Life Technologies, Inc.), supplemented with 1% penicillin/streptomycin (Gibco). Cell lines were regularly tested and verified to be mycoplasma negative using MycoAlert Detection Kit (Lonza) or via PCR.

Primary mouse PDAC cell lines were established from pancreatic tumours in respective GEMMs as described previously<sup>12</sup>. HY15549 and HY19636 cells were established from female KPC mice (*p48-cre<sup>+</sup>* (*p48* is also known as *Ptf1a*) *Kras<sup>LSL-G12D/+</sup>*, *Trp53<sup>lox/+</sup>*)<sup>43</sup> that were fully backcrossed into a C57BL/6 background. The other cell lines were derived from the following mice<sup>10,12</sup>: *Atg5<sup>+/-</sup>* KPC cells, *Pdx1-cre<sup>+</sup>*, *Kras<sup>LSL-G12D/+</sup>*, *Trp53<sup>lox/+</sup>*, *Atg5<sup>+/-</sup>* mice; *Atg5<sup>-/-</sup>* KPC cells, *Pdx1-cre<sup>+</sup>*, *Kras<sup>LSL-G12D/+</sup>*, *Trp53<sup>lox/+</sup>*, *Atg5<sup>lox/lox</sup>* mice; AY6284 cells, *p48-cre<sup>+</sup>*, *Kras<sup>LSL-G12D/+</sup>*, *Trp53<sup>lox/+</sup>*, *Rosa<sup>LSL-rtTA</sup>*, mSt-ATG4B(C74A) mice<sup>12</sup>. All mouse PDAC cells were maintained in DMEM (Corning) supplemented with 10% FBS (Atlanta Biologicals, S11550H) and 1% penicillin/streptomycin (Gibco).

Cells were grown in a humidified incubator with 5% CO<sub>2</sub> at 37 °C. Cultures were routinely verified to be negative for mycoplasma. Cell lines were authenticated by fingerprinting, and low passage cultures were carefully maintained in a central lab cell bank.

### Constructs

GFP–NBR1 and GFP–NBR1 dUBA were provided by J. Debnath. LysoTag TMEM192-mRFP-3xHA (TMRHA) was generated by subcloning the cDNA of TMEM192 (Origene) together with monomeric red fluorescent protein (mRFP) and 3× haemagglutinin (HA) tag into the NheI and EcoRI sites of pLJM1 lentiviral vector. HLA-A-TurboID-Flag (HLA-A-TrID) was generated by subcloning the cDNA of HLA-A (Addgene plasmid, 85162) into the EcoRI and NotI sites of the TurboID pLVX vector (gift from R. Zoncu). pMXs GFP–LC3–RFP was a gift from N. Mizushima (Addgene, plasmid 117413). For Dox-inducible expression of ATG4B(C74A), mTurquoise2 or mStrawberry was fused to ATG4B(C74A) and inserted into either pSLIK-Hygro (used for in vitro studies) (Addgene, plasmid 25737) or pINDUCER20 (used for in vivo studies) (Addgene, plasmid 44012), using the Gateway Cloning system (Thermo Fisher Science). For the generation of OVA-expressing cells, the cOVA fragment was cloned from pCI-neo-cOVA (Addgene, plasmid 25097), fused with 2A peptide and mStrawberry sequences using NEBuilder HiFi DNA Assembly Cloning Kit (New England BioLabs) according to manufacturer's instruction, and inserted into the EcoRI and SalI sites of pBabe-zeo (Addgene, plasmid 1766) to generate pBabe-cOVA-2A-mStrawberry. Stable cOVA expression was confirmed and monitored with mStrawberry fluorescence.

### shRNAs

shRNA vectors (pLKO.1 puro) were obtained from the Sigma MISSION TRC shRNA library. The sequences and RNAi Consortium clone IDs for the shRNAs used are as follows:

shATG3#1 (human): 5'-GATGTGACCATTGACCATATT-3' (TRCN0000148120); shATG3#2 (human): 5'-GCTGTGCTTCCAACAATAGAA-3' (TRCN0000147381); shATG7#1 (human): 5'-CCCAGCTATTGGAACACGTGA-3' (TRCN0000007587); shATG7#2 (human): 5'-GCCTGCTGAGGAGCTCTCCAT-3' (TRCN0000007584); shNRB1#1 (human): 5'-GCTTCATAGTTATTTGGCATT-3' (TRCN0000123159); shNRB1#2 (human): 5'-GCAGCATTTGTGGATGAGAAT-3' (TRCN0000123160); shNRB1#3 (human): 5'-GCCAGGAACCAAGTTATCAA-3' (TRCN0000123161);

shFIP200#1 (human): 5'-GCACTCTTTAACACATTCTTT-3' (TRCN0000013523); shFIP200#2 (human): 5'-GCTGTGAATGAGTTTGAATA-3' (TRCN0000013524); shATG14#1 (human): 5'-CCATAGAAGCTTGGTCA TGT-3' (TRCN0000144080); shATG14#2 (human): 5'-GATCAATTAC AACCCTGCAT-3' (TRCN0000145367); shAtg3#1 (mouse): 5'-CATATCA CAACACAGGTATTA-3' (TRCN0000247440); shAtg3#2 (mouse): 5'-GTACATCACTTACGACAAATA-3' (TRCN0000247442); shAtg7#1 (mouse): 5'-TTCTGTACCGTTTCGATAATG-3' (TRCN0000305991); shAtg7#2 (mouse): 5'-GCCAACATCCCTGGATACAAG-3' (TRCN0000375444); shAtg7#3 (mouse): 5'-TCTTACCCTGCTCCATCAAGA-3' (TRCN0000375421); shB2m (mouse): 5'-CCAGTTTCTAATATGCTATAC-3' (TRCN0000295705); shAtg13#1 (mouse): 5'-TGAAGTCTCTTCTCGCT ATTA-3' (TRCN0000277121); shAtg13#2 (mouse): 5'-GACATACCTTTCG CCATGTTT-3' (TRCN0000176029); shUlk1#1 (mouse): 5'-CGCTTCTTT CTGGACAAACAA-3' (TRCN0000319764); shUlk1#2 (mouse): 5'-CGCTT CTTTCTGGACAAACAA-3' (TRCN0000028768); shGFP: 5'-TGCCCGACA ACCACTACCTGA-3' (TRCN0000072186). shGFP and shScr (Addgene, plasmid #17920) were used as controls.

### siRNAs

The short interfering RNAs (siRNAs) used in this study are: siNC (silencer negative control #1 siRNA, Thermo, 13778030), siRubicon #1 (si04762, Ambion), and siRubicon #2 (si04763, Ambion). Cells were transfected with siRNAs using Lipofectamine RNAi Max Transfection Reagent (Life Technologies).

### Retroviral and lentiviral transduction

For the transfection of pMXs GFP-LC3-RFP<sup>44</sup> and pBabe-cOVA-2A-mStrawberry, retrovirus was produced by co-transfection of HEK293FT cells with a retroviral vector and the packaging plasmids pHIT60 and VSVG at a 0.5:0.25:0.25 ratio. For the transfection of lentiviral vectors (pSLIK-hygro, pINDUCER20 and pLKO.1-puro), lentivirus was produced by co-transfection of HEK293FT cells with a lentiviral vector and the packaging plasmids psPAX2 (Addgene, plasmid 12260) and pMD2.G (Addgene, plasmid 12259) at a 0.5:0.25:0.25 ratio. Transfection was performed using Lipofectamine 3000 Reagent (Thermo Fisher Scientific) according to manufacturer's instructions. The viral supernatant was collected 48 h after transfection, filtered through a 0.45-µm filter, and used for infection together with Polybrene reagent (EMD Millipore). Cells expressing Dox-inducible constructs were sorted for positive fluorescent expression after Dox treatment to select inducible cells, and then sorted for no fluorescent expression after Dox withdrawal to remove cells with leaking expression, as described previously<sup>8</sup>.

### Immunofluorescence

Human cell lines were cultured for two days on coverslips coated with fibronectin. After two PBS washes, cells were fixed and permeabilized with paraformaldehyde for 15 min at room temperature or ice-cold methanol for 5 min at -20 °C. Paraformaldehyde-fixed cells were permeabilized with 0.1% saponin. Samples were then blocked with 5% normal goat serum for 15 min at room temperature before incubation with primary antibodies (Supplementary Table 1) overnight at 4 °C. After washing three times with PBS, cells were incubated in secondary antibody at room temperature for 20 min. Slides were mounted on glass slides using DAPI Fluoromount-G (0100-20, SouthernBiotech) and imaged on a Zeiss Laser Scanning Microscope (LSM) 710 using a 63× objective. Image processing and quantification were performed using ImageJ.

### Lysosome immunoprecipitation

For lysosome immunoprecipitation experiments, HPDE and human PDAC cell lines stably expressing TMRHA were ruptured and intact lysosomes from 1–2 mg of total protein per sample was immunoprecipitated using HA-conjugated Dynabeads as previously described<sup>45,46</sup>.

### Poly-ubiquitin UBA affinity capture

Cells were washed twice with ice-cold PBS and lysed in lysis buffer (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 1% NP-40, 10 mM *N*-ethylmaleimide) supplemented fresh with protease inhibitor cocktail. Lysates were incubated at 4 °C for 15 min and clarified by centrifugation at 20,000  $\times$ g at 4 °C for 15 min. Samples were quantified by BCA Protein Assay Kit and diluted to 1 mg ml<sup>-1</sup> with dilution buffer (50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 5 mM EDTA, protease inhibitor). Approximately 1–1.5 mg of protein lysates were incubated with 50  $\mu$ l of Ubiquitin 1 Tandem UBA Agarose (BostonBiochem AM-130) overnight at 4 °C. Samples were then washed three times in High Salt Wash Buffer (50 mM Tris-HCl pH 7.5, 250 mM NaCl, 0.5% NP-40) and once with 10 mM Tris-HCl, pH 7.5. Samples were eluted by adding Laemlli buffer and incubating at 65 °C for 15–20 min.

For DUB digestions, affinity captured material was washed once with DUB digestion buffer (50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 20 mM dithiothreitol). Liquid was removed and beads were resuspended in 20 ml of DUB digestion buffer and 5 mg of USP2 catalytic domain (Usp2cc; BostonBiochem E-506) for 1 h with gentle shaking at 30 °C. Beads were washed once in high-salt wash buffer and once with 10 mM Tris-HCl, pH 7.5, before eluting affinity captured material by adding Laemlli buffer and incubating at 65 °C for 15 min.

### Proximity biotinylation

Human PDAC cells stably expressing HLA-A–TrID were cultured in DMEM supplemented with 10% dialysed FBS (DMEM + dFBS) for 48 h. Cells were incubated with 10  $\mu$ M biotin (Sigma) and incubated at 37 °C for 30 min. Media was replaced with DMEM + dFBS and incubated at 37 °C for a further 2–3 h. For negative controls, we omitted exogenous biotin. Cells were then washed twice in ice-cold PBS and lysed in ice-cold lysis buffer (1% Triton X-100, 130 mM NaCl, 2.5 mM MgCl<sub>2</sub>, 2 mM EGTA, 25 mM HEPES pH 7.4, supplemented fresh with protease inhibitor cocktail) for 30 min on ice. Samples were clarified by centrifugation at 13,300 rpm for 10 min at 4 °C. Protein content was measured using BCA Protein Assay Kit (Life Technologies 23227). Then, 1–2 mg of protein lysates were incubated with 50  $\mu$ l of Dynabeads MyOne Streptavidin C1 (Life Technologies) overnight. The beads were washed twice in wash buffer 1 (2% SDS in dH<sub>2</sub>O), once in wash buffer 2 (0.1% deoxycholate, 1% Triton X-100, 500 mM NaCl, 1 mM EDTA, and 50 mM HEPES, pH 7.5), once in wash buffer 3 (250 mM LiCl, 0.5% NP-40, 0.5% deoxycholate, 1 mM EDTA, and 10 mM Tris, pH 8.1), and twice in wash buffer 4 (50 mM Tris, pH 7.4, and 50 mM NaCl). Washes were performed at room temperature for 5 min with gentle agitation. Samples were eluted in Laemlli buffer and boiled at 95 °C. Quantification of biotinylation activity was measured as intensity of each lane in the +biotin immune-precipitation condition divided by the corresponding intensity of the ligase expression band in the input. The +biotin ratios were then normalized to the –biotin control ratio and to background.

### Organoid culture

Normal pancreatic organoids from mice were established as previously described<sup>25,47</sup> with slight modifications. In brief, pancreas was harvested from female C57BL/6 mice (4–10 weeks of age), mechanically minced with scissors, and digested with 1 mg ml<sup>-1</sup> collagenase P (Sigma), 4 mg ml<sup>-1</sup> Dispase II (Sigma), 10 mM HEPES (Thermo Fisher Scientific), 1% FBS, 1 mg ml<sup>-1</sup> Trypsin Inhibitor from Soybean (Sigma) in DMEM for 20 min at 37 °C. Digested tissues were embedded in Matrigel (Corning, 356231) and cultured in Advanced DMEM (Thermo Fisher Scientific) supplemented with 200 mM L-glutamine (Thermo Fisher Scientific), 10 mM HEPES (Thermo Fisher Scientific), 1% Antibiotic-Antimycotic (Thermo Fisher Scientific), 500 ng ml<sup>-1</sup> recombinant murine rSpondin-1 (Peprotech), 50 ng ml<sup>-1</sup> recombinant murine EGF (Peprotech), 100 ng ml<sup>-1</sup> recombinant murine noggin (Peprotech), 1  $\mu$ M jagged-1

(188-204) (AnaSpec, AS-61298), 100  $\mu$ g ml<sup>-1</sup> trypsin inhibitor from soybean (Sigma), and 10  $\mu$ M Y-27632 (Enzo). Two days before analysis, culture media was replaced with DMEM supplemented with 10% FBS. PDAC organoids were generated by embedding mouse PDAC cell lines in Matrigel and were grown in DMEM with 10% FBS. For the dissociation of organoids for further assays, organoids were digested with TrypLE (Thermo Fisher, A1217701) solution at a 2 $\times$  concentration for 20 min at 37 °C, followed by filtration through a 40- $\mu$ m nylon strainer.

### Quantitative proteomics

Quantitative mass spectrometry-based proteomics was performed as previously described<sup>48,49</sup> based on the SL-TMT workflow<sup>50</sup>. In brief, cells were lysed in a lysis buffer (200 mM HEPES pH 8.5, 8 M urea, 1 $\times$  Complete Protease Inhibitor Cocktail (Roche), 1 $\times$  PhosStop (Roche)) and homogenized by passing through a 21-gauge needle. Lysates were collected by centrifuging at 20,000g for 5 min at 4 °C, followed by disulfide bond reduction with 5 mM dithiothreitol at 37 °C for 25 min and alkylation with 10 mM iodoacetamide at room temperature for 30 min in the dark. Chloroform–methanol precipitation of protein was performed, followed by protease digestion in HEPES buffer (200 mM, pH 8.5). Each sample containing 100  $\mu$ g protein was digested at a 1:100 protease-to-protein ratio with LysC protease at room temperature overnight, followed by digestion with trypsin at 37 °C for 6 h. Approximately 50  $\mu$ g of peptides from each sample was labelled with 100  $\mu$ g TMT reagent which were dissolved in anhydrous acetonitrile to achieve a final concentration of 30% (v/v). TMT-labelled samples was acidified, vacuum centrifuged to near dryness and subjected to C18 SPE (Sep-Pak, Waters). Samples were subjected to basic pH reversed-phase HPLC. Data were obtained with Orbitrap Fusion mass spectrometer (Thermo Fisher Scientific) coupled with a Proxeon EASY-nLC 1000 LC pump (Thermo Fisher Scientific). Peptide separation was done using a custom Accucore C18 resin (2.6  $\mu$ m, 100 Å, Thermo Fisher Scientific) column for 3 h using a gradient of 6–30% acetonitrile in 0.125% formic acid with a flow rate of 300 nl min<sup>-1</sup>. All analysis involved an MS<sup>3</sup>-based TMT method as previously mentioned<sup>51</sup> and mass spectra were processed as described earlier<sup>48</sup>.

### Tumour cell and OT-I cell co-culture experiment

Mouse PDAC cells stably expressing OVA and carrying Dox-inducible mTurquoise2-ATG4B(C74A) were grown as organoids, treated with or without Dox (1  $\mu$ g ml<sup>-1</sup>) for 96 h. Organoids were dissociated into single cells, which were then incubated with either anti-H-2K<sup>b</sup>-SIINFEKL antibody (clone 25-D1.16, BioXCell, BE0207) or isotype control (clone MOPC-21, BioXCell, BE0083) at 100  $\mu$ g ml<sup>-1</sup> for 30 min at 4 °C. Total splenocytes were harvested from OT-I mice and CD8<sup>+</sup> T cells were enriched using Dynabeads Untouched Mouse CD8 Cells (Invitrogen, 11417D) following manufacturer's instructions. Isolated CD8<sup>+</sup> T cells were labelled with 10  $\mu$ M CFSE (BioLegend) for 10 min at room temperature in the dark, washed three times with RPMI-1640 supplemented with 10% FBS. Ten thousand PDAC cells and forty thousand CD8<sup>+</sup> cells were seeded in 96-well plates and cultured in 100  $\mu$ l 50% DMEM and 50% RPMI-1640 supplemented with 10% FBS, 10 ng ml<sup>-1</sup> recombinant murine IL-2 (Peprotech), 27.5  $\mu$ M 2-mercaptoethanol (Gibco), and 100  $\mu$ g ml<sup>-1</sup> of respective antibodies. After 48 h, CD8<sup>+</sup> T cells were harvested and stained with anti-CD8a antibody (AF647, clone 53-6.7, BioLegend) and DAPI, and proliferation was analysed by CFSE dilution using flow cytometry. After removal of CD8<sup>+</sup> T cells, the viability of remaining PDAC cells was measured by CellTiter-Glo (Promega).

### Clonogenic assay

For 2D clonogenic assays<sup>9</sup>, cells were plated in 6-well plates at 300 cells per well in 2 ml DMEM supplemented with 10% FBS. After 7 days, colonies were fixed with 80% methanol, stained with 0.2% crystal violet, and counted. For 3D clonogenic assay, single cells were sorted directly

# Article

into 384-well round bottom Ultra-Low Attachment plates (Corning 3830) and grown in DMEM supplemented with 10% FBS and 2% Matrigel (Corning 356231). After 10 days, the number of wells with spheroids was counted.

## Mice

Female 8–10-week-old C57BL/6 mice or NCr nude (B6NTac, Taconic) were used for allograft experiments. OT-I transgenic mice (003831) and *Batf3*<sup>-/-</sup> mice (013755) were purchased from Jackson Laboratory. All mice were bred and maintained in the animal facility of the New York University School of Medicine. All animal procedures were approved by the New York University School of Medicine Institutional Animal Care and Use Committee (IACUC) under protocol numbers IA16-00507 and IA16-01331.

## Mouse experiments

Orthotopic and intrasplenic injections of PDAC cells were performed as described previously<sup>31,52</sup>. In brief, mice were anaesthetized by an intraperitoneal injection of ketamine and xylazine. A small incision was made on the upper left quadrant of the abdomen, and either the pancreas or the spleen was externalized. For orthotopic injection, cells were suspended in 20 µl of Matrigel (Corning 356231):HBSS (1:1) solution and injected into the pancreatic tail with insulin syringes (29-gauge needle, BD 324702). Approximately  $1 \times 10^4$  HY15549 cells or  $2 \times 10^4$  HY19636 cells were injected unless otherwise indicated. For intrasplenic injection,  $1 \times 10^6$  cells were suspended in 100 µl HBSS and then drawn into an insulin syringe (28-gauge needle, BD 329461), which was pre-loaded with 200 µl HBSS. The externalized spleen was divided by ligating clips (Teleflex, 002200), and cells were injected into the hemispleen. After injection, splenic vein was ligated with ligating clips (Teleflex, 001200) at the hilum of the spleen, and then the hemispleen was removed. After the procedures, the peritoneum was closed with a 3-0 VICRYL VIOLET suture (Ethicon, J311H), and the skin was closed using the BD AutoClip Wound Closing System (BD). For some experiments, mice were fed with Dox-containing diet (625 mg kg<sup>-1</sup>) for the indicated period. Mice were euthanized at the indicated time points and tumours or the liver were collected after trans-cardiac perfusion with PBS.

For CD8<sup>+</sup> T cell depletion, mice received intraperitoneal injection of anti-mouse CD8a antibody (200 µg, clone 53-6.7, BE0004-1) or isotype control (200 µg, clone 2A3, BE0089). For immune checkpoint blockade experiments, mice received intraperitoneal injection of anti-mouse PD-1 antibody (200 µg, clone RMP1-14, BE0146) and anti-mouse CTLA-4 antibody (200 µg, BE0131), or rat IgG2a isotype control (200 µg, clone 2A3, BE0089) and polyclonal Syrian hamster IgG (200 µg, BE0087). All antibodies used in *in vivo* experiments were obtained from BioXCell. For chloroquine treatment, mice received intraperitoneal injection of chloroquine solution in PBS at 60 mg kg<sup>-1</sup> or PBS every day starting day 4 or 5 after tumour cell implantation. Mice were randomly assigned to specific treatment groups at the beginning of treatment.

All experiments were carried out in a clean conventional facility, where mice were housed in pre-packaged disposable irradiated cages, and fed with irradiated diet and acidified water. Microisolator cages were located on ventilated racks. No tumours in the mice exceeded IACUC-defined maximum diameters of >2 cm. Sample sizes were determined based on our preliminary experiments and no sample size calculation was done. Blinding was not performed as the investigator needed to know the treatment groups in order to perform study. Tumour weights (an objective measurement) were measured only at the study endpoints after mice were euthanized and tumours were harvested.

**Flow cytometry.** For surface and intracellular MHC-I staining of human cell lines, cells were stained with Alexa Fluor 488 anti-human HLA-A,

-B, -C antibody (BioLegend, clone W6/32) at a 1:75 dilution for 45 min at 4 °C in the dark and washed with PBS plus 2% FBS and 2 mM EDTA (FACS buffer). Cells were then fixed and permeabilized before staining with phycoerythrin (PE)-conjugated anti-human HLA-A, -B, -C antibody (BioLegend, clone W6/32) at a 1:75 dilution for 45 min at room temperature and washed with FACS buffer. For cell-surface molecule staining of mouse cells, single-cell suspensions were prepared as described above. Cells were washed with FCM buffer (HBSS containing 1% FBS, 1 mM EDTA, and 10 mM HEPES) and stained with antibodies (Supplementary Table 1) at 4 °C in the dark for 20 min. Dead cells were depleted by DAPI staining.

For the immunophenotyping of tumours, tissues were mechanically minced with scissors, and then digested in DMEM containing 1 mg ml<sup>-1</sup> collagenase IV (Gibco), 100 µg ml<sup>-1</sup> DNase I (Roche), 1% FBS, 10 mM HEPES, 2% Antibiotic-Antimycotic (Thermo Fisher Scientific) for 40 min at 37 °C in the dark with gentle agitation every 10 min. Digested tissues were then washed twice in DMEM containing 10% FBS, filtered through a 40-µm nylon mesh strainer (Corning). Cells were suspended in ACK Lysing Buffer (Thermo Fisher Scientific), incubated for 10 min at 4 °C in the dark to remove red blood cells. Cells were washed twice in FCM buffer and counted. Cells were stained with Zombie Aqua Fixable Viability Kit (BioLegend) and blocked with anti-mouse CD16/CD32 antibody (Mouse BD Fc Block, clone 2.4G2, BD Bioscience). One million cells were incubated with appropriate antibodies (Supplementary Table 1) diluted in FCM buffer at 4 °C in the dark for 40 min. Cells were then washed twice with FCM buffer and analysed or further fixed in 2% PL solution (PBS containing 0.1 M L-lysine (Sigma) and 2% paraformaldehyde (Ted Pella))<sup>53</sup> at 4 °C in the dark overnight. Intracellular staining was performed using Foxp3/Transcription Factor Staining Buffer Set (eBioscience). Cells were analysed on a BD LSR Fortessa or a BD LSR-II UV and analysed by FlowJo software (FlowJo, LLC, v.10.4). Gating strategies were described in Extended Data Fig. 7k and Supplementary Table 2.

## Cell sorting

Mouse PDAC cells (HY15549) expressing the GFP-LC3-RFP reporter were grown as organoids for 8 days, collected and dissociated into single cells as mentioned above. Cells were sorted on MoFlo XDP (Beckman Coulter) based on the GFP/RFP ratio as described in detail elsewhere<sup>54,55</sup>. The purity of cells after sorting was confirmed by post-sort analyses and was usually >90%.

## Histology and immunohistochemistry

For formalin-fixed paraffin-embedded (FFPE) sections, tissues were fixed in 10% buffered formalin, and embedded in paraffin. Five-micrometre-thick FFPE sections were used for H&E staining. For immunohistochemical staining, FFPE sections were deparaffinized, rehydrated and incubated in boiling 10 mM pH 6.0 citrate buffer for 20 min for antigen retrieval. To visualize the GFP-LC3-RFP reporter signal or mStrawberry/mSt-ATG4B(C74A) expression in PDAC cells, tissues were fixed in 2% PL solution at 4 °C overnight, incubated in PBS containing 30% sucrose, and embedded in Tissu-Tek OCT compound (Sakura Finetek) on dry ice<sup>53</sup>. For immunohistochemical staining, 7-µm frozen sections were washed with Tris-buffered saline (pH 7.4) containing 0.05% Tween-20 (TBS-T), blocked with TBS-T containing 5% goat serum for 1 hr at room temperature, and stained with primary antibodies (Supplementary Table 1) at 4 °C overnight, followed by incubation with fluorescent-conjugated secondary antibodies (1:400, Supplementary Table 1) at room temperature for 1 h. Sections were counter stained with 4 µg ml<sup>-1</sup> Hoechst 33342 (Thermo Fisher) and mounted in ProLong Diamond Antifade Mountant (Thermo Fisher Scientific, P36970). Fluorescent images were obtained with Leica DM6 and analysed using LAS X software (v.2.0.0.14332.2). Bright light images were obtained with a Leica DM2000 bright-field microscope.

For the measurement of the GFP/RFP signal ratio, frozen sections as prepared above were washed with TBS-T three times and coverslips were mounted with ProLong Diamond Antifade Mountant. Slides were dried at 4 °C overnight. Fluorescence images were obtained from at least four random fields per tumour with a 20× objective lens. Greyscale, raw-image files (16-bit) were analysed using ImageJ Software to obtain mean intensities of GFP and RFP signals.

Primary human PDAC specimens (Supplementary Table 3) were generated under Institutional Review Board (IRB)-approved protocol 18-25787 at UCSF. Sections were co-stained with antibodies against CK-19 (1:300) and HLA-A,B (1:300) (Supplementary Table 1). An average of 28 ducts per sample from 9 patient specimens was imaged based on expression of CK-19 to ensure analysis of tumour epithelia. The corresponding MHC-I localization per CK-19 positive ductal structure was classified as intracellular if greater than 50% of the staining displayed a non-membrane, punctate distribution.

### Quantitative reverse-transcription PCR

Total RNA was extracted using TRIzol (Thermo Fisher Scientific) and PureLink RNA Mini Kit (Thermo Fisher Scientific), and then reversed transcribed using Superscript Vilo IV (Thermo Fisher Scientific) with oligo-dT primers. Quantitative PCR was performed with SYBR Green Supermix (Bio-rad) on the CFX96 real-time PCR machine (Bio-Rad). The quantity of mRNA was calculated using the  $\Delta C_t$  method and normalized by the *GAPDH*, *Actb* or *36B4* (also known as *RplpO*) genes for humans or mice, respectively. Sequences for qPCR primers are listed in Supplementary Table 4.

### Transcriptome analysis

RNA-seq libraries were prepared using the Illumina TruSeq Stranded mRNA library preparation kit according to manufacturer's instructions by the NYU Genome Technology Center. The libraries were pooled and sequenced as 50-base, single-end reads on an Illumina HiSeq 4000 using high output mode (v4 chemistry). The raw fastq reads were aligned to mm10 mouse reference genome using STAR aligner<sup>56</sup>. Fastq Screen was used to check for any contaminations in the samples and Picard RnaSeqMetrics was used to obtain the metrics of all aligned RNA-seq reads. featureCounts was used to quantify the gene expression levels<sup>57</sup>. FPKM (fragments per kilobase of transcript per million fragments mapped) data were used as input for gene set enrichment analysis (GSEA)<sup>58</sup>.

### Immunoblotting

Cells were lysed in cell lysis buffer (Cell Signaling Technology, 9803) supplemented with protease inhibitor (Thermo Fisher Scientific, A32953) and phosphatase inhibitor (PhosSTOP, Sigma Aldrich, 04906837001). Proteins were separated on 12% or 4–20% Mini-PROTEAN TGX Precast Protein Gels by SDS PAGE electrophoresis and transferred to onto PVDF membranes (IPVH00010, EMD Millipore) or nitrocellulose membranes (GE Healthcare Amersham Protran NC Rolls, Fisher, 10600000). Membranes were blocked in 5% non-fat dry milk (Blotting-Grade Blocker, BioRad, 1706404) dissolved in TBS-T for 1 h and incubated overnight at 4 °C with primary antibodies (Supplementary Table 1). Membranes were washed three times with TBS-T and incubated with secondary antibodies for 1 h at room temperature: anti-rabbit IgG, horseradish peroxidase (HRP)-linked (Cell Signaling Technology, 7074), anti-mouse IgG, HRP-linked (Cell Signaling Technology, 7076). Images were obtained by chemiluminescence (Bio-rad 1705061) using a ChemiDoc (Bio-Rad).

### Statistical analysis

Statistical analysis was performed using Graphpad Prism 7.0 (Graph-Pad). Results are expressed as mean  $\pm$  s.d. unless otherwise indicated. For each box-and-whisker plot, centre line is the median and whiskers represent the minimum and maximum values.

### Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

### Data availability

RNA-seq data have been deposited to the Gene Expression Omnibus (GEO) data repository with accession number GSE145766. Source Data are provided for all experiments. Other data that support the findings of this study are available on request from the corresponding author upon reasonable request.

42. Liu, N., Furukawa, T., Kobari, M. & Tsao, M. S. Comparative phenotypic studies of duct epithelial cell lines derived from normal human pancreas and pancreatic carcinoma. *Am. J. Pathol.* **153**, 263–269 (1998).
43. Bardeesy, N. et al. Both p16<sup>INK4a</sup> and the p19<sup>Arf</sup>-p53 pathway constrain progression of pancreatic adenocarcinoma in the mouse. *Proc. Natl Acad. Sci. USA* **103**, 5947–5952 (2006).
44. Kaizuka, T. et al. An autophagic flux probe that releases an internal control. *Mol. Cell* **64**, 835–849 (2016).
45. Zoncu, R. et al. mTORC1 senses lysosomal amino acids through an inside-out mechanism that requires the vacuolar H<sup>+</sup>-ATPase. *Science* **334**, 678–683 (2011).
46. Abu-Remaileh, M. et al. Lysosomal metabolomics reveals V-ATPase- and mTOR-dependent regulation of amino acid efflux from lysosomes. *Science* **358**, 807–813 (2017).
47. Matsuura, T. et al. Organoid-based ex vivo reconstitution of Kras-driven pancreatic ductal carcinogenesis. *Carcinogenesis* bgz122 (2019).
48. Paulo, J. A., Gaun, A. & Gygi, S. P. Global analysis of protein expression and phosphorylation levels in nicotine-treated pancreatic stellate cells. *J. Proteome Res.* **14**, 4246–4256 (2015).
49. Biancur, D. E. et al. Compensatory metabolic networks in pancreatic cancers upon perturbation of glutamine metabolism. *Nat. Commun.* **8**, 15965 (2017).
50. Navarrete-Perea, J., Yu, Q., Gygi, S. P. & Paulo, J. A. Streamlined tandem mass tag (SL-TMT) protocol: an efficient strategy for quantitative (phospho)proteome profiling using tandem mass tag-synchronous precursor selection-MS3. *J. Proteome Res.* **17**, 2226–2236 (2018).
51. McAllister, G. C. et al. MultiNotch MS3 enables accurate, sensitive, and multiplexed detection of differential expression across cancer cell line proteomes. *Anal. Chem.* **86**, 7150–7158 (2014).
52. Soares, K. C. et al. A preclinical murine model of hepatic metastases. *J. Vis. Exp.* **51677**, 51677 (2014).
53. Hirata, Y. et al. CD150<sup>high</sup> bone marrow Tregs maintain hematopoietic stem cell quiescence and immune privilege via adenosine. *Cell Stem Cell* **22**, 445–453 (2018).
54. Gump, J. M. et al. Autophagy variation within a cell population determines cell fate through selective degradation of Fap-1. *Nat. Cell Biol.* **16**, 47–54 (2014).
55. Gump, J. M. & Thorburn, A. Sorting cells for basal and induced autophagic flux by quantitative ratiometric flow cytometry. *Autophagy* **10**, 1327–1334 (2014).
56. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
57. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
58. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).

**Acknowledgements** This work was supported by National Cancer Institute Grants R01CA157490, R01CA188048, P01CA117969, R35CA232124; ACS Research Scholar Grant RSG-13-298-01-TBG; NIH grant R01GM095567; and the Lustgarten Foundation, and SU2C to A.C.K. R.M.P. is the Nadia's Gift Foundation Innovator of the Damon Runyon Cancer Research Foundation (DRR-46-17) and is additionally supported by an NIH Director's New Innovator Award (1DP2CA216364) and the Pancreatic Cancer Action Network Career Development Award. K.Y. was supported by the Uehara Memorial Foundation Research Fellowship. A.V. is supported by a National Science Foundation Graduate Research Fellowship. D.E.B. is supported by a Ruth L. Kirschstein Institutional National Research Service Award, T32 CA009161 (Levy), and the NCI Predoctoral to Postdoctoral Fellow Transition Award (F99/K00) F99 CA245822. M.K. was supported by the Uehara Memorial Foundation Research Fellowship and Postdoctoral Fellowship for Research Abroad (Japan Society for the Promotion of Science). S.J.P. was supported by American Cancer Society grant 132942-PF-18-215-01-TBG. R.S.B. is a Merck Fellow of the Damon Runyon Cancer Research Foundation (DRG-2348-18). We thank H. Ying for providing the HY19636 and HY15549 cells. We thank the New York University (NYU) Langone Health Experimental Pathology Laboratory, Flow Cytometry Core, and Genome Technology Center, each supported in part by the Cancer Center Support grant P30CA016087 at the Laura and Isaac Perlmutter Cancer Center. We thank Z. Dewan for the technical assists on the staining of frozen sections. We acknowledge the UCSF Parnassus Flow Cytometry Core (PFCC) supported in part by the DRC Center Grant P30DK063720 for assistance generating flow cytometry data. We thank J. Olzmann for advice on ubiquitylation experiments. We thank all members of Perera, Kimmelman, Debnath and Pacold laboratories for suggestions.

**Author contributions** K.Y. and A.V. performed most experiments and wrote the manuscript. J.Y. assisted with cloning and performed the proximity biotinylation and ubiquitylation experiments. D.E.B. and A.S.W.S. assisted with animal studies. S.G. performed immunofluorescence and analysis of patient PDAC specimens. M.K. assisted with the analysis



# Article

of flow cytometry data and RNA-seq data. S.M. assisted with immunoblotting and preparing shRNAs. E.Y.L. and S.J.P. cloned fluorescent constructs. K.W.W. and G.E.K. provided PDAC patient specimens and analysis. J.D. provided GFP-NBR1 and GFP-NBR1 dUBA constructs. R.S.B. assisted with transcriptome data analysis. J.D.M. and J.A.P. performed proteomics analysis. D.T.F. provided intellectual feedback and support. R.M.P. and A.C.K. conceived the project, supervised the research and wrote and edited the paper.

**Competing interests** A.C.K. has financial interests in Vescor Therapeutics, LLC. A.C.K. is an inventor on patents pertaining to KRAS-regulated metabolic pathways, redox control pathways in pancreatic cancer, targeting GOT1 as a therapeutic approach, and the autophagic control of iron metabolism. A.C.K. is on the SAB of Rafael/Cornerstone Pharmaceuticals. A.C.K.

is a consultant for Deciphera. J.D. is on the Scientific Advisory Board of Vescor Therapeutics, LLC. The other authors declare no competing interests.

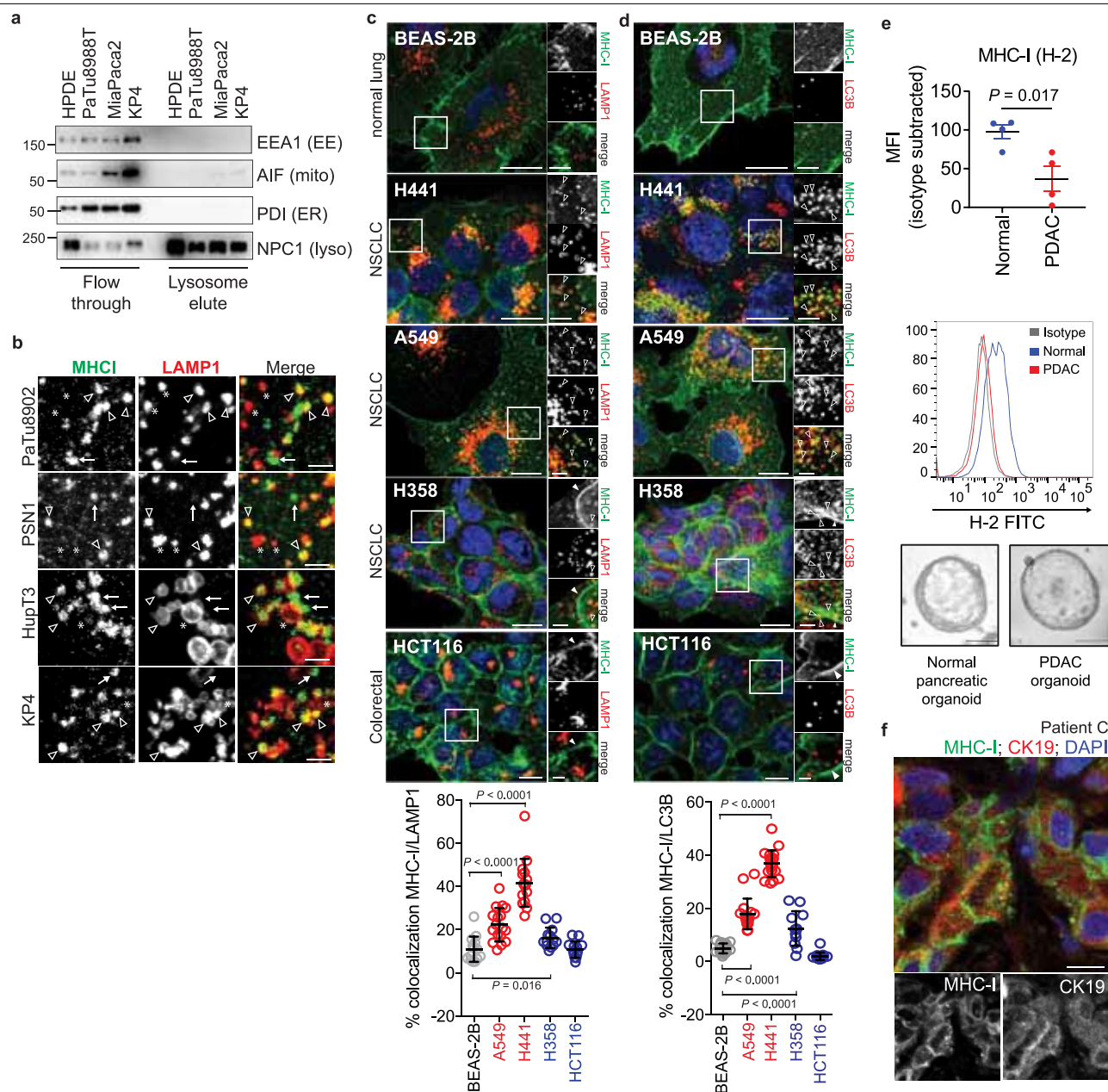
## Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41586-020-2229-5>.

**Correspondence and requests for materials** should be addressed to R.M.P. or A.C.K.

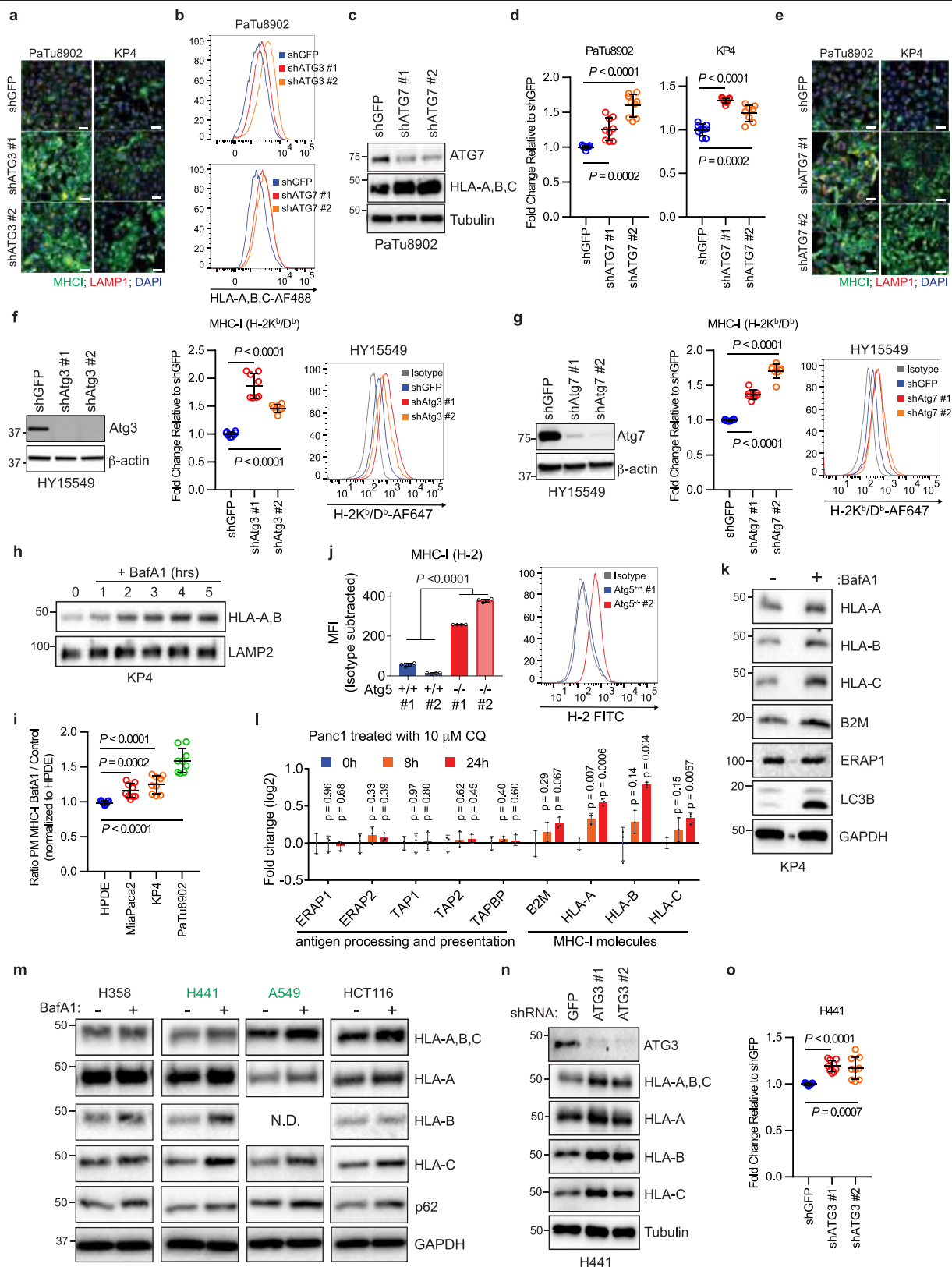
**Peer review information** *Nature* thanks Christian Münz, Andrew Thorburn, Robert H. Vonderheide and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>.



**Extended Data Fig. 1 | Heterogeneous distribution of MHC-I in KRAS-mutant cancers.** **a**, Immuno-isolation of intact lysosomes from HPDE and PDAC cell lines showing absence of non-lysosome markers as indicated. EE, early endosome; ER, endoplasmic reticulum; lyso, lysosome; mito, mitochondria. **b**, High-power images showing MHC-I-positive/LAMP1-positive (arrowheads), MHC-I-positive/LAMP1-negative (arrows) and MHC-I-negative/LAMP1-positive (asterisk) puncta. Scale bars, 5  $\mu$ m. **c, d**, Top, localization of MHC-I (green) relative to LAMP1-positive (red) lysosomes (BEAS-2B,  $n = 14$ ; A549,  $n = 17$ ; H441,  $n = 15$ ; H358,  $n = 13$ ; HCT116,  $n = 13$ ) (c) or LC3B-positive (red) autophagosomes (BEAS-2B,  $n = 18$ ; A549,  $n = 17$ ; H441,  $n = 20$ ; H358,  $n = 12$ ; HCT116,  $n = 15$ ) (d) in the indicated cell lines. Bottom, quantification of percentage co-localization. Cell lines indicated in red show significantly

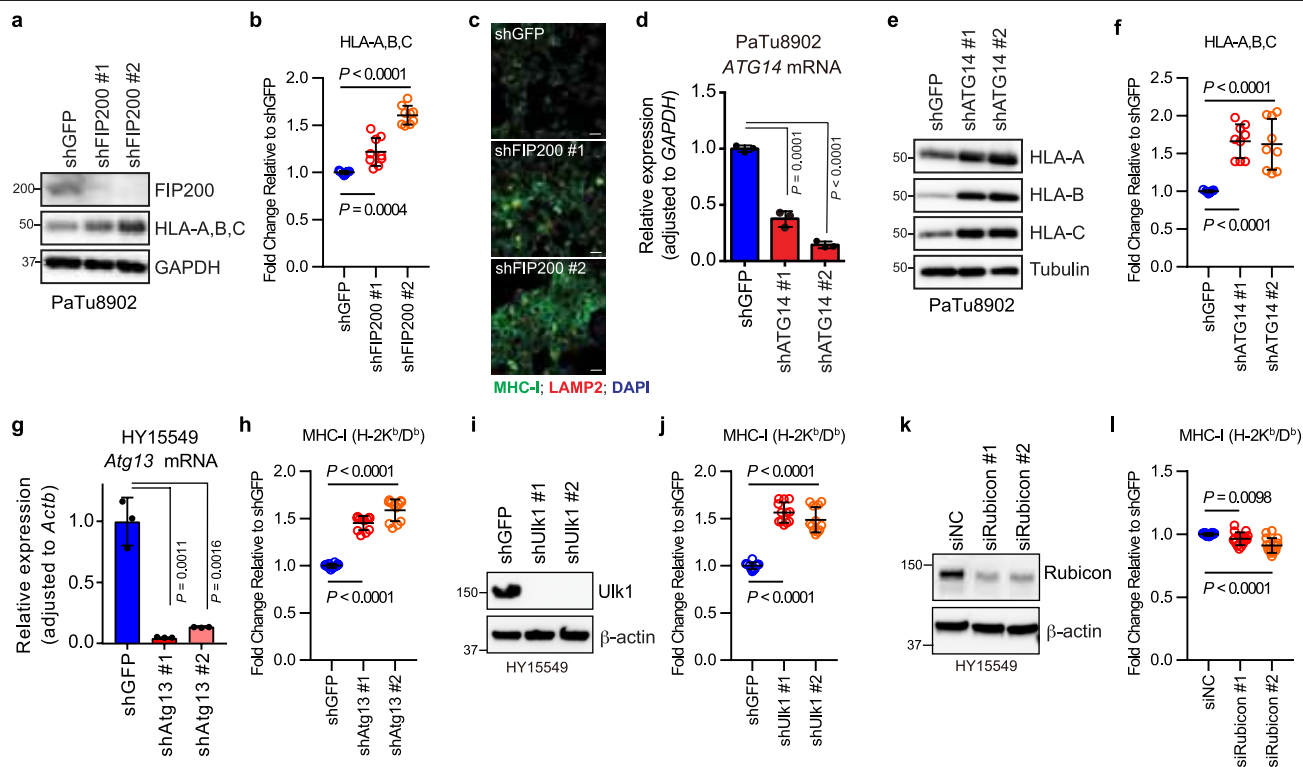
increased co-localization relative to BEAS-2B cells, and cell lines indicated in blue show a modest increase (H358) or no difference (HCT116). Data are mean  $\pm$  s.d. (c, d). Scale bars, 20  $\mu$ m and 10  $\mu$ m (inset). **e**, Flow cytometry-based analysis of surface MHC-I (H-2) in mouse normal pancreas (C57Bl/6) and mouse PDAC cells grown as organoids. Top, isotype-subtracted geometric MFI. Each dot represents different mice/lines ( $n = 4$ ). Data are mean  $\pm$  s.e.m. Middle, representative flow cytometry plots. Bottom, representative images of organoids. **f**, Immunofluorescent staining images from a patient in Fig. 1g showing intracellular localization of MHC-I (green) in CK19-positive (red) ducts. Scale bar, 20  $\mu$ m. A representative of at least two independent experiments is shown in a and e.  $P$  values determined by unpaired two-tailed  $t$ -tests (c–e). See Supplementary Fig. 1 for gel source data.



Extended Data Fig. 2 | See next page for caption.

**Extended Data Fig. 2 | Inhibition of autophagy and lysosomes restores MHC-I levels and plasma membrane localization.** **a**, Immunofluorescence staining of MHC-I after shRNA-mediated ATG3 knockdown. Scale bars, 50  $\mu$ m. **b**, Representative flow cytometry plots for PaTu8902 cells after knockdown of ATG3 (related to Fig. 2b) and ATG7 (see also **d**). Representative plots from **d** and Fig. 2b are shown. **c**, Effect of ATG7 knockdown on MHC-I (HLA-A, -B, -C) expression in PaTu8902 cells. **d**, Flow cytometry-based quantification of plasma membrane levels of MHC-I (HLA-A, -B, -C) after ATG7 knockdown ( $n = 9$  replicates from three independent experiments). **e**, Immunofluorescence staining of MHC-I following ATG7 knockdown. Scale bars, 50  $\mu$ m. **f, g**, Surface MHC-I levels after knockdown of ATG3 (**f**) or ATG7 (**g**) in mouse PDAC cells. Left, knockdown efficiency was confirmed by immunoblots. Middle, cell surface levels of MHC-I (H-2K<sup>b</sup>/D<sup>b</sup>) measured by flow cytometry ( $n = 8$  replicates from two independent experiments). Right, representative flow cytometry plots are shown. **h**, Treatment of KP4 cells with 150 nM BafA1 for the indicated times causes an increase in levels of HLA-A, -B. **i**, Flow cytometry-based quantification of plasma membrane MHC-I in the indicated cell lines after

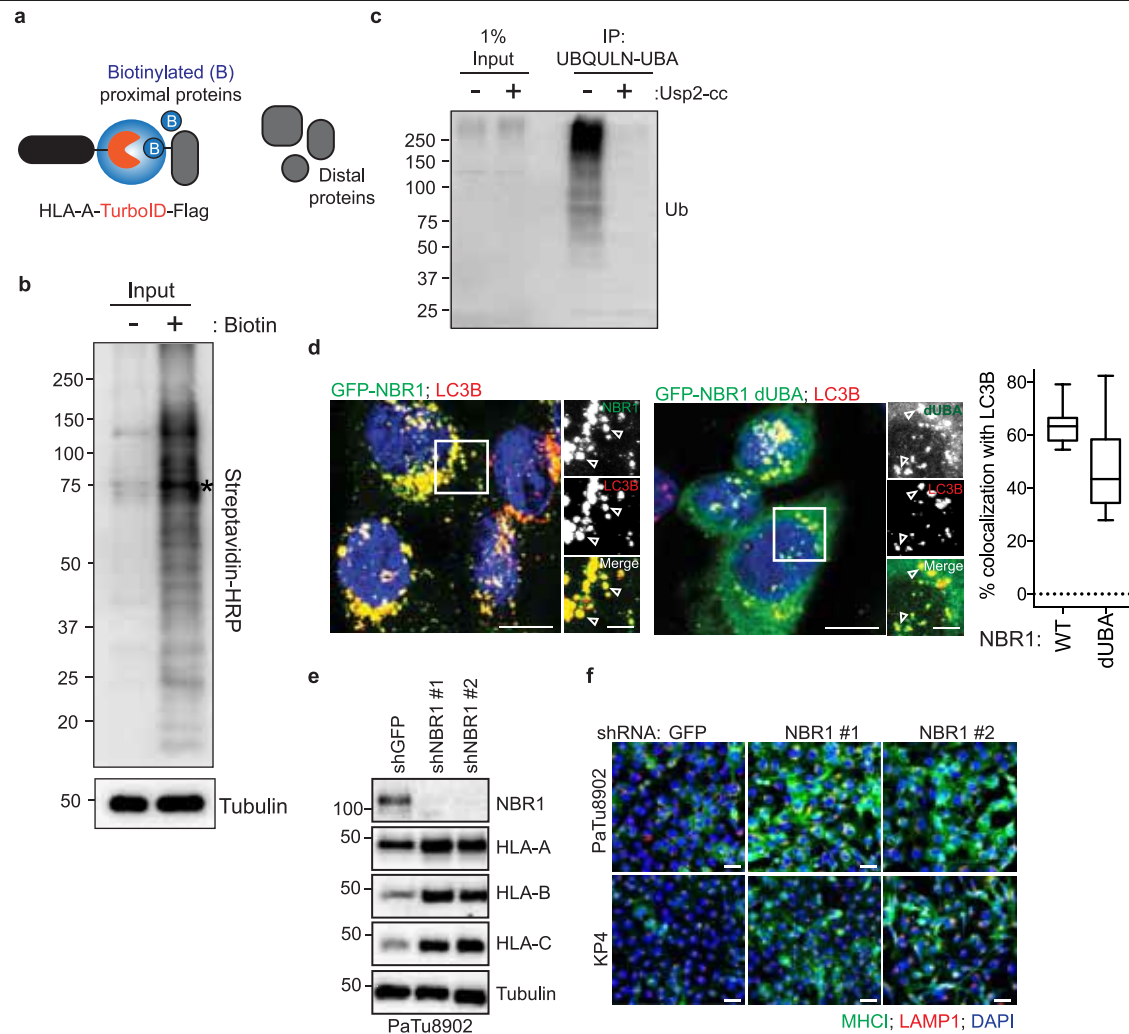
treatment with BafA1 for 16 h ( $n = 9$  replicates from three independent experiments). **j**, Surface MHC-I (H-2) levels measured by flow cytometry. Mouse PDAC organoids were established from *Atg5*<sup>+/+</sup> and *Atg5*<sup>-/-</sup> KPC cells.  $n = 4$  biological replicates. Data are representative of three independent experiments. Right, representative flow cytometry plots. **k**, Effect of BafA1 treatment on the expression levels of antigen presentation machinery. **l**, Quantitative proteomics analysis of Panc1 cells that were treated with chloroquine (10  $\mu$ M) for the indicated periods.  $n = 3$  biological replicates. **m**, Effect of BafA1 treatment on expression levels of MHC-I in the indicated cell lines. Cell lines denoted in green show a significant change across all HLA isoforms after BafA1 treatment. **n, o**, Effect of ATG3 knockdown in H441 cells on total MHC-I (**n**) and plasma membrane MHC-I (**o**) as measured by flow cytometry-based quantification ( $n = 9$  replicates from three independent experiments). A representative of at least two independent experiments is shown in **a–c, e–h, j, k, m, n**. Data are mean  $\pm$  s.d. *P* values were determined by unpaired two-tailed *t*-tests. See Supplementary Fig. 1 for gel source data.



**Extended Data Fig. 3 | Inhibition of macroautophagy, but not LAP/LANDO, restores MHC-I levels.** Knockdown mediated by shRNA (a–j) or siRNA (k, l) of FIP200, ATG14, ATG13 and ULK1, but not RUBICON, increased MHC-I levels in PDAC cells. **a, d, g, i, k.** Knockdown efficiency was confirmed by immunoblot (a, i, k) and qPCR (d, g). Data are mean  $\pm$  s.d. from three biological replicates per group (d, g). **a, e.** Whole-cell abundance of MHC-I was assessed by immunoblot. **c.** Immunofluorescence staining of MHC-I (green) and LAMP2 (red). Scale bars,

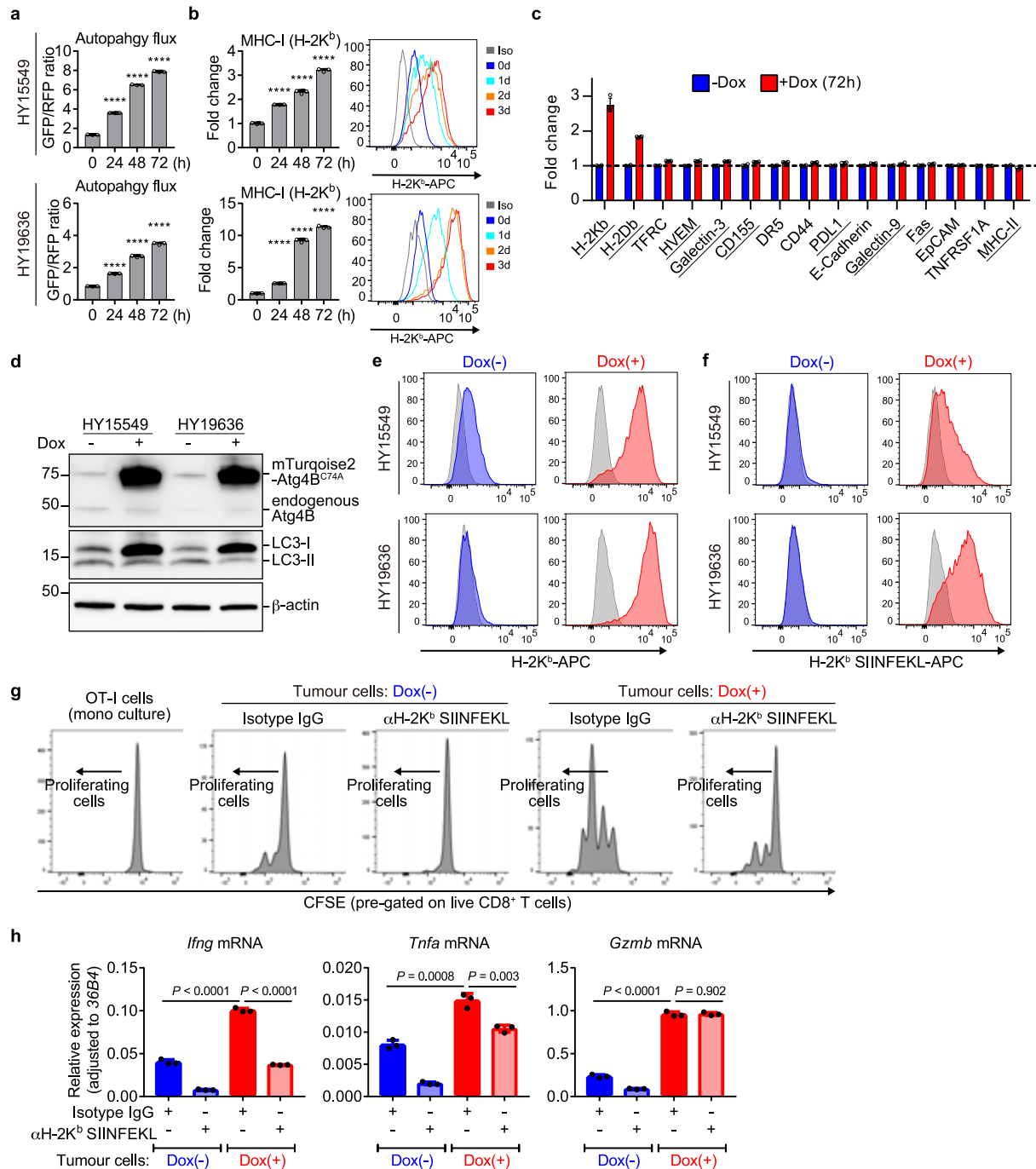
50  $\mu$ m. **b, f, h, j, l.** Cell-surface MHC-I levels were measured by flow cytometry (b, f, n = 9; h, j, n = 12; l, n = 16). Data are pooled from at least three independent experiments. Data are mean  $\pm$  s.d. **a–f,** PaTu8902 cells (human). **g–l,** HY15549 cells (mouse). A representative of at least two independent experiments is shown in a, c, e, i and k. *P* values determined by unpaired two-tailed *t*-tests. See Supplementary Fig. 1 for gel source data.





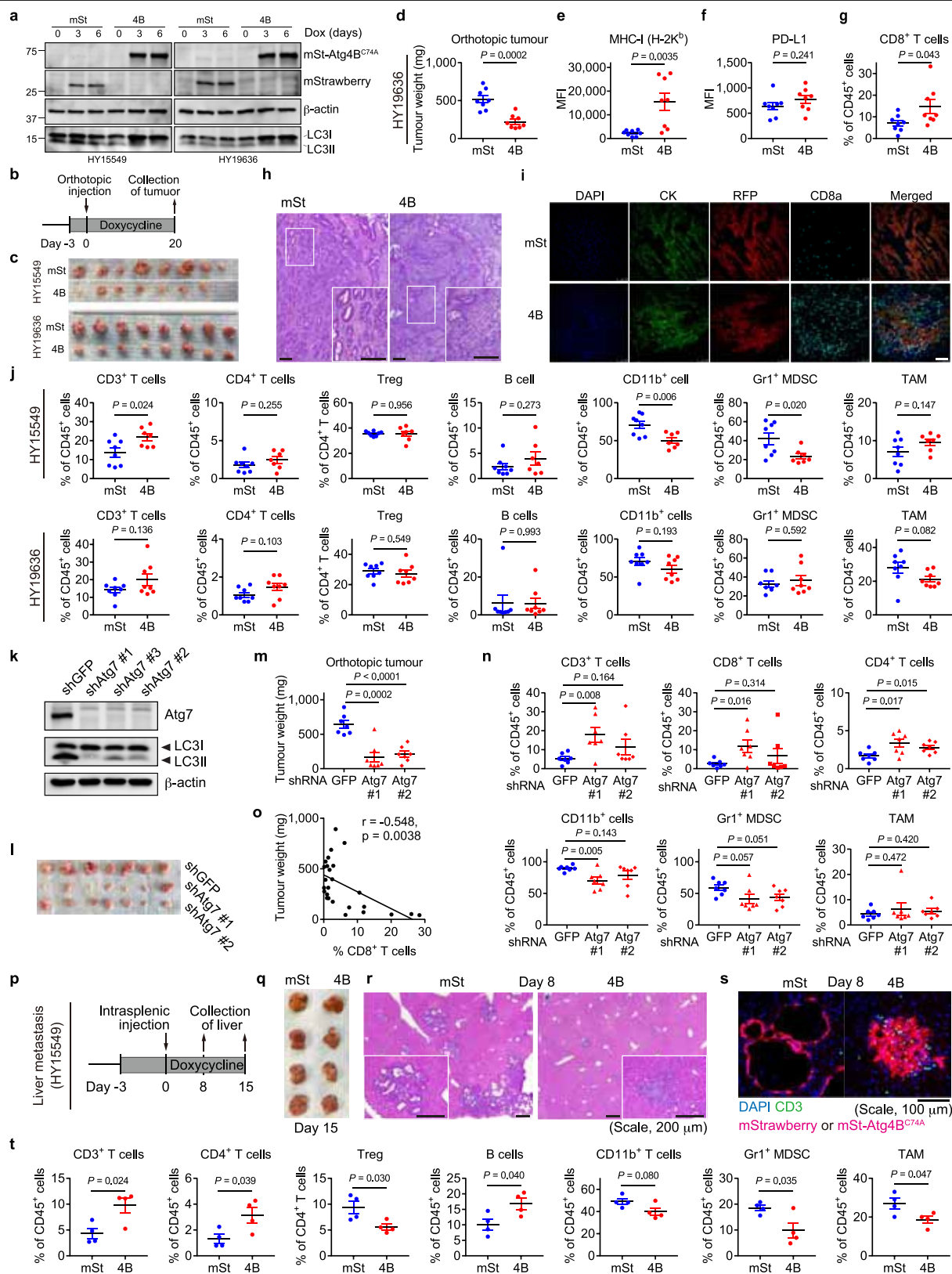
**Extended Data Fig. 4 | The UBA domain of NBR1 is required for interaction with MHC-I.** **a**, Proximity-dependent biotinylation catalysed by HLA-A-TrID. After addition of biotin, TurboID catalyses the formation of biotin-5'-AMP anhydride, which enables covalent tagging of endogenous proteins with biotin within a few nanometres of the ligase. Related to Fig. 2c. **b**, HLA-A-TrID was stably expressed in KP4 cells. Cells were treated with 10  $\mu$ M of exogenous biotin for 30 min. After labelling, cells were lysed and biotinylated proteins were enriched with streptavidin conjugated beads. Biotinylated proteins were detected using streptavidin-HRP (**b**) or with antibodies against the indicated proteins (see Fig. 2c). Asterisks indicates ligase self-biotinylation. **c**, Endogenous ubiquitylated proteins were affinity captured from PaTu8902 cells with UBQLN1 UBA conjugated beads. Treatment of affinity captured

samples for 1 h with purified Usp2-cc (+) to induce deubiquitylation leads to loss of ubiquitylation. Related to Fig. 2e. **d**, PaTu8902 cells stably expressing wild-type NBR1 (GFP-NBR1,  $n = 19$  fields) or lacking the UBA domain (GFP-NBR1 dUBA,  $n = 16$  fields) were co-stained for endogenous LC3B. Graph shows quantification of the percentage co-localization. Box-and-whisker plots as in Fig. 1. Scale bars, 20  $\mu$ m (inset 10  $\mu$ m). Related to Fig. 2f. **e**, Effect of NBR1 knockdown on respective HLA-A, -B and -C levels in PaTu8902 cells. Note that blotting images for NBR1 and tubulin are the same as in Fig. 2g. **f**, Immunofluorescence staining of MHC-I after NBR1 knockdown. Scale bars, 50  $\mu$ m. A representative of at least three independent experiments is shown in **b**, **c**, **e**, **f**. See Supplementary Fig. 1 for gel source data.



**Extended Data Fig. 5 | Autophagy inhibition restores MHC-I expression, leading to enhanced anti-tumour T cell response in vitro.** **a, b**, Autophagy flux (**a**) and cell-surface MHC-I levels (**b**) in PDAC cells measured by flow cytometry. Mouse PDAC cells expressing the GFP-LC3-RFP reporter and Dox-inducible mTurquoise2-ATG4B(C74A) were grown as organoids for 8 days and treated with Dox ( $1 \mu\text{g ml}^{-1}$ ) for the indicated hours. **a**, Autophagy flux represented by GFP/RFP ratio. Note that increased GFP/RFP ratio indicates reduced autophagy flux. **b**, Cell surface MHC-I (H-2K<sup>b</sup>) levels. Representative flow cytometry plots are shown. Data are mean  $\pm$  s.d.  $n = 3$  biological replicates. Data are representative of at least four independent experiments. **c**, Fold changes of respective molecules on the cell surface quantified by flow cytometry. HY15549 cells expressing Dox-inducible mTurquoise2-tagged ATG4B(C74A) were grown as organoids for 8 days and treated with or without Dox ( $1 \mu\text{g ml}^{-1}$ ) for 72 h. Positive surface expression of each molecule was confirmed using respective isotype controls. Molecules found in immunological synapses are underlined. TFRC, transferrin receptor. Data are mean  $\pm$  s.d.  $n = 4$  biological replicates. Representative data from two

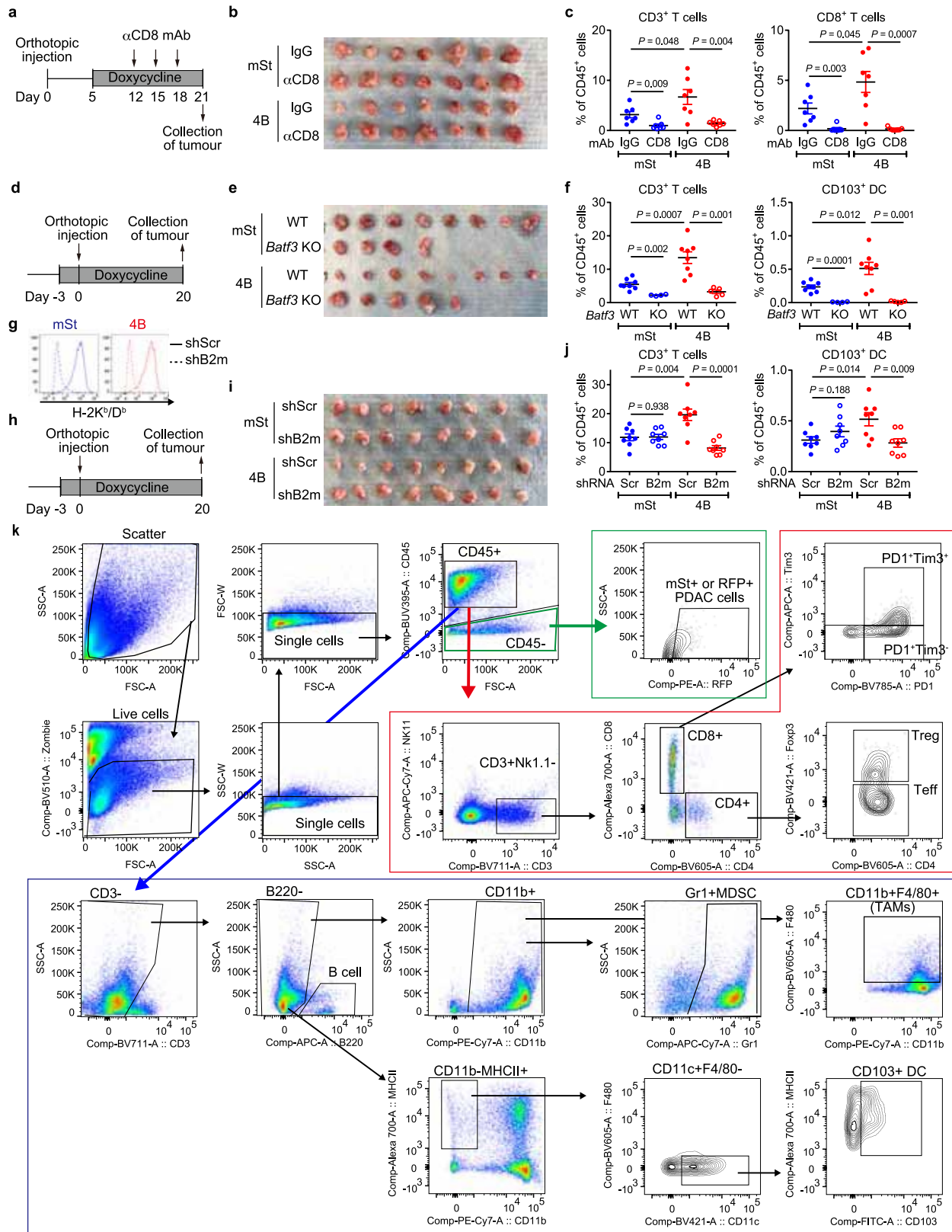
independent experiments are shown. **d-f**, Mouse PDAC cells expressing OVA and carrying Dox-inducible mTurquoise2-ATG4B(C74A) were grown as organoids and treated with or without Dox ( $1 \mu\text{g ml}^{-1}$ ) for 96 h. Related to Fig. 3a, b. **d**, Autophagy inhibition was confirmed by immunoblot. mTurquoise2-ATG4B(C74A) or endogenous ATG4B were detected by anti-ATG4B antibody. **e, f**, Flow cytometry plots for H-2K<sup>b</sup> (**e**) and H-2K<sup>b</sup> SIINFEKL (**f**). Representative plots from Fig. 3a, b are shown. Grey, isotype control. **g**, Representative flow cytometry plots of the OT-I cells co-cultured with mouse PDAC cells from Fig. 3c. **h**, Quantitative reverse transcription PCR (qRT-PCR) analysis of OT-I cells that were co-cultured with PDAC cells for 48 h. Related to Fig. 3c. Data are mean  $\pm$  s.d.  $n = 3$  biological replicates. For **g** and **h**, Dox(+) or Dox(-) indicates that PDAC cells were grown with or without Dox ( $1 \mu\text{g ml}^{-1}$ ) before co-culture. Dox was not added in co-culture. A representative of at least three independent experiments is shown in **d-g**. \*\*\*\* $P < 0.0001$ , unpaired two-tailed  $t$ -tests. See Supplementary Fig. 1 for gel source data.



**Extended Data Fig. 6** | See next page for caption.

**Extended Data Fig. 6 | Autophagy inhibition modulates anti-tumour immunity in both orthotopic tumours and liver metastasis.** **a**, Immunoblots showing autophagy inhibition in mSt-ATG4B(C74A)-expressing cells. Mouse PDAC cells carrying Dox-inducible mSt or 4B were treated with Dox ( $1 \mu\text{g ml}^{-1}$ ) for the indicated days. mSt or mSt-ATG4B(C74A) was detected by anti-RFP antibody. A representative of two independent experiments is shown. **b–j**, Related to Fig. 3e–h. Mouse PDAC cells shown in **a** were orthotopically transplanted into syngeneic (C57BL/6) mice. HY15549 cells (mSt,  $n = 8$ ; 4B,  $n = 7$ ) and HY19636 cells ( $n = 8$  per group) were injected. **b**, Study design. **c**, Images of tumours at end point. **d–g**, HY19636 tumour weight (**d**), cell surface MHC-I levels (**e**) and PD-L1 levels (**f**) on PDAC cells and tumour-infiltrating CD8<sup>+</sup> T cells (**g**) measured by flow cytometry. **h, i**, Representative H&E staining (**h**) and immunofluorescent staining (**i**) of HY15549 tumours (mSt,  $n = 8$ ; 4B,  $n = 7$ ). Scale bars,  $100 \mu\text{m}$ . **j**, Quantification of tumour-infiltrating immune cells by flow cytometry (HY15549,  $n = 8$  and  $7$ ; HY19636,  $n = 8$  per group). Gating strategies are shown in Extended Data Fig. 7k and Supplementary Table 2. MDSC, myeloid-derived suppressor cells; Treg, T regulatory cells; TAM, tumour-associated macrophages. **k–o**, Autophagy-inhibition by shRNA-mediated ATG7 knockdown elicits similar anti-tumour T cell responses.

**k**, Immunoblots for ATG7, LC3 and  $\beta$ -actin in PDAC cells (HY15549) expressing shRNAs against GFP or ATG7. A representative of at least two independent experiments is shown. **l–o**, Mouse PDAC cells shown in **k** were orthotopically transplanted into syngeneic mice ( $n = 7$  per group). **l**, Images of tumours collected on day 22. **m**, Tumour weight. **n**, Tumour-infiltrating immune cells as measured by flow cytometry. **o**, Correlation between CD8<sup>+</sup> T cell frequency among CD45<sup>+</sup> cells and tumour weight. **p–t**, Related to Fig. 3i–l. Autophagy inhibition modulates anti-tumour immunity in metastatic tumours in the liver. Mouse PDAC cells (HY15549) carrying Dox-inducible mSt or 4B were injected into the spleen of syngeneic (C57BL/6) mice that were pre-fed with a Dox-containing diet ( $n = 4$  per group). PDAC cells were pre-treated with Dox ( $1 \mu\text{g ml}^{-1}$ ) for 7 days before injection. **p**, Study design. **q**, Images of the liver. **r, s**, Representative images of H&E staining (**r**) and immunofluorescent staining (**s**) ( $n = 4$  per group). Scale bars,  $200 \mu\text{m}$  (**r**) and  $100 \mu\text{m}$  (**s**). **t**, Quantification of immune cells in the liver metastasis as measured by flow cytometry. Data are mean  $\pm$  s.e.m.  $n$  indicates individual mice.  $P$  values were determined by unpaired two-tailed  $t$ -tests (**d–g, j, m, n, t**) and Pearson correlation analysis (**o**). See Supplementary Fig. 1 for gel source data.

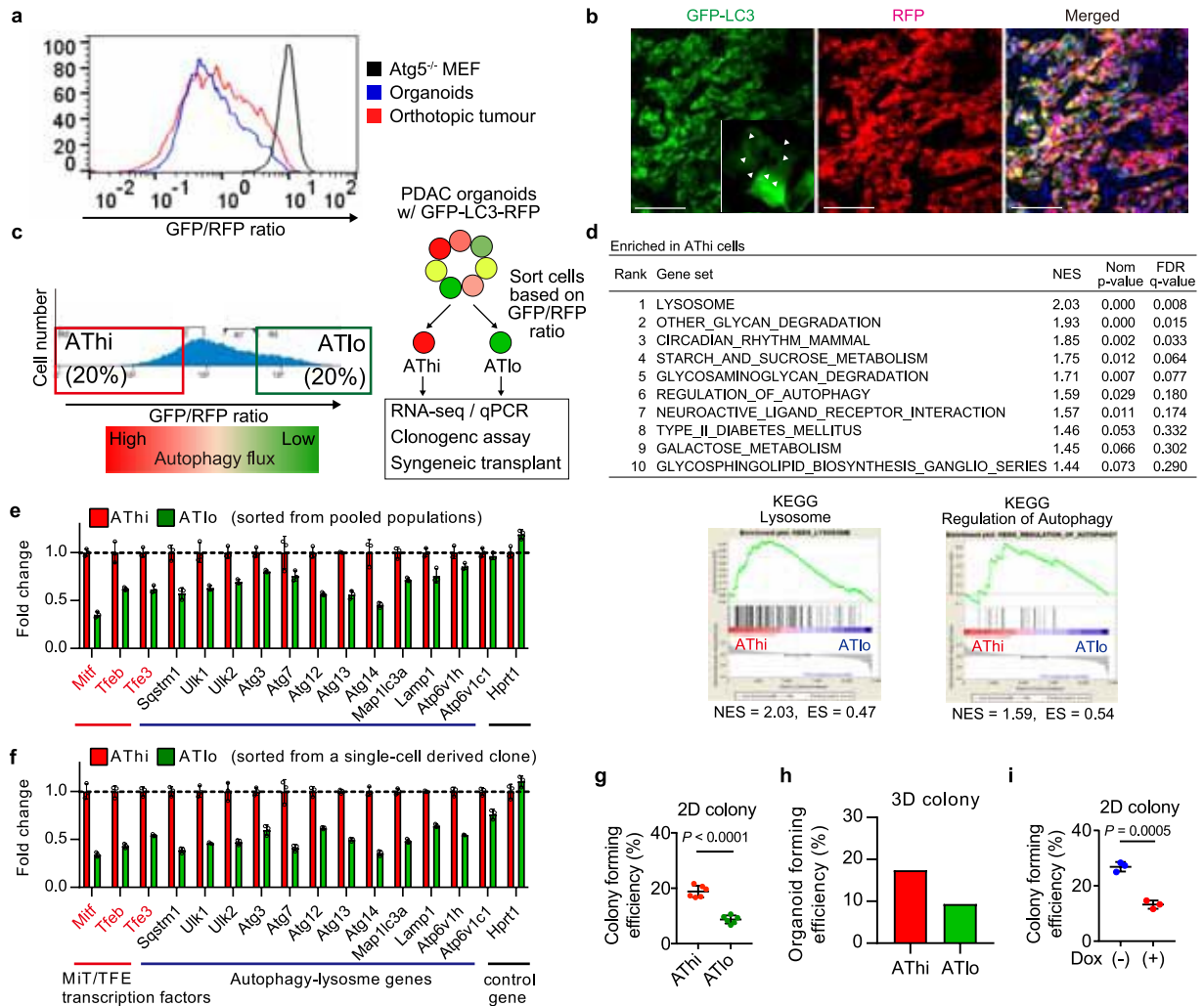


**Extended Data Fig. 7** | See next page for caption.



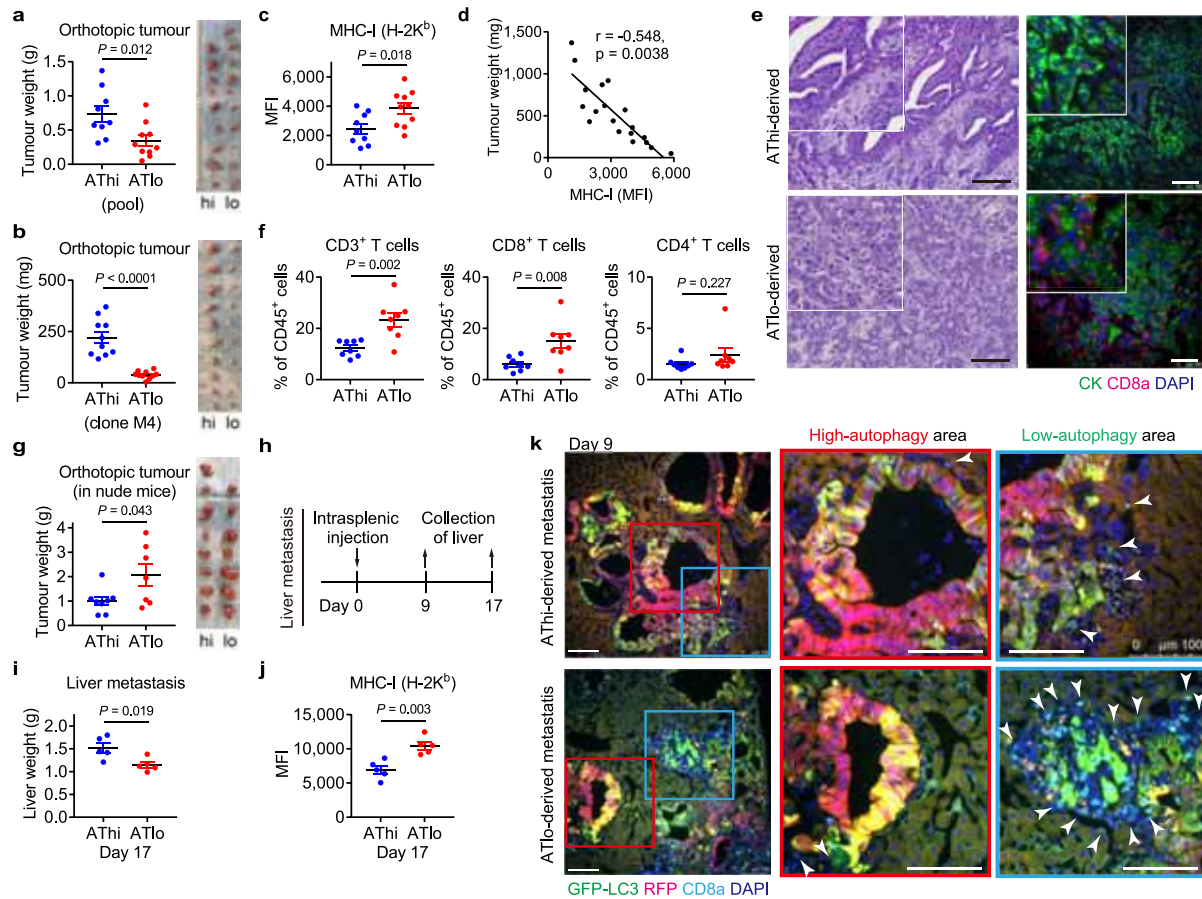
**Extended Data Fig. 7 | Tumour regression after autophagy inhibition is rescued by depletion of CD8<sup>+</sup> T cells or ablation of cell surface MHC-I. a–c,** Related to Fig. 3m. HY15549 cells with Dox-inducible mSt or 4B were orthotopically injected into C57BL/6 mice and fed with Dox-containing diet starting on day 5, and then received intraperitoneal injection of anti-CD8 or isotype control IgG ( $n = 7$  per group). **a**, Study design. **b**, Images of tumours. **c**, Tumour-infiltrating leukocytes as quantified by flow cytometry. **d–f**, Related to Fig. 3n, o. HY15549 cells with Dox-inducible mSt or 4B were orthotopically injected into C57BL/6 mice (WT) or *Batf3*<sup>-/-</sup> mice (KO) ( $n = 8, 4, 8$  and  $5$ ). **d**, Study design. **e**, Images of tumours. **f**, Quantification of tumour-infiltrating leukocytes by flow cytometry. **g–j**, Related to Fig. 3p–r. HY15549 cells carrying

Dox-inducible mSt or 4B were stably transfected with lentiviral vectors expressing control shRNA (shScr, solid line) or shRNA against B2m (shB2m, dashed line). **g**, Cell surface MHC-I as measured by flow cytometry. Cells were treated with IFN $\gamma$  ( $200 \text{ U ml}^{-1}$ ) for 24 h before flow cytometry analysis. Representative data from three independent experiments are shown. **h–j**, Cells shown in **g** ( $4 \times 10^4$  cells) were orthotopically transplanted into syngeneic (C57BL/6) mice that were pre-fed with Dox diet ( $n = 8$  per group). **h**, Study design. **i**, Images of tumours. **j**, Tumour-infiltrating leukocytes as quantified by flow cytometry. **k**, Gating strategies for flow cytometry analysis of tumours used in this study. See also Supplementary Table 2. Data are mean  $\pm$  s.e.m.  $n$  indicates individual mice.  $P$  values determined by unpaired two-tailed  $t$ -tests.



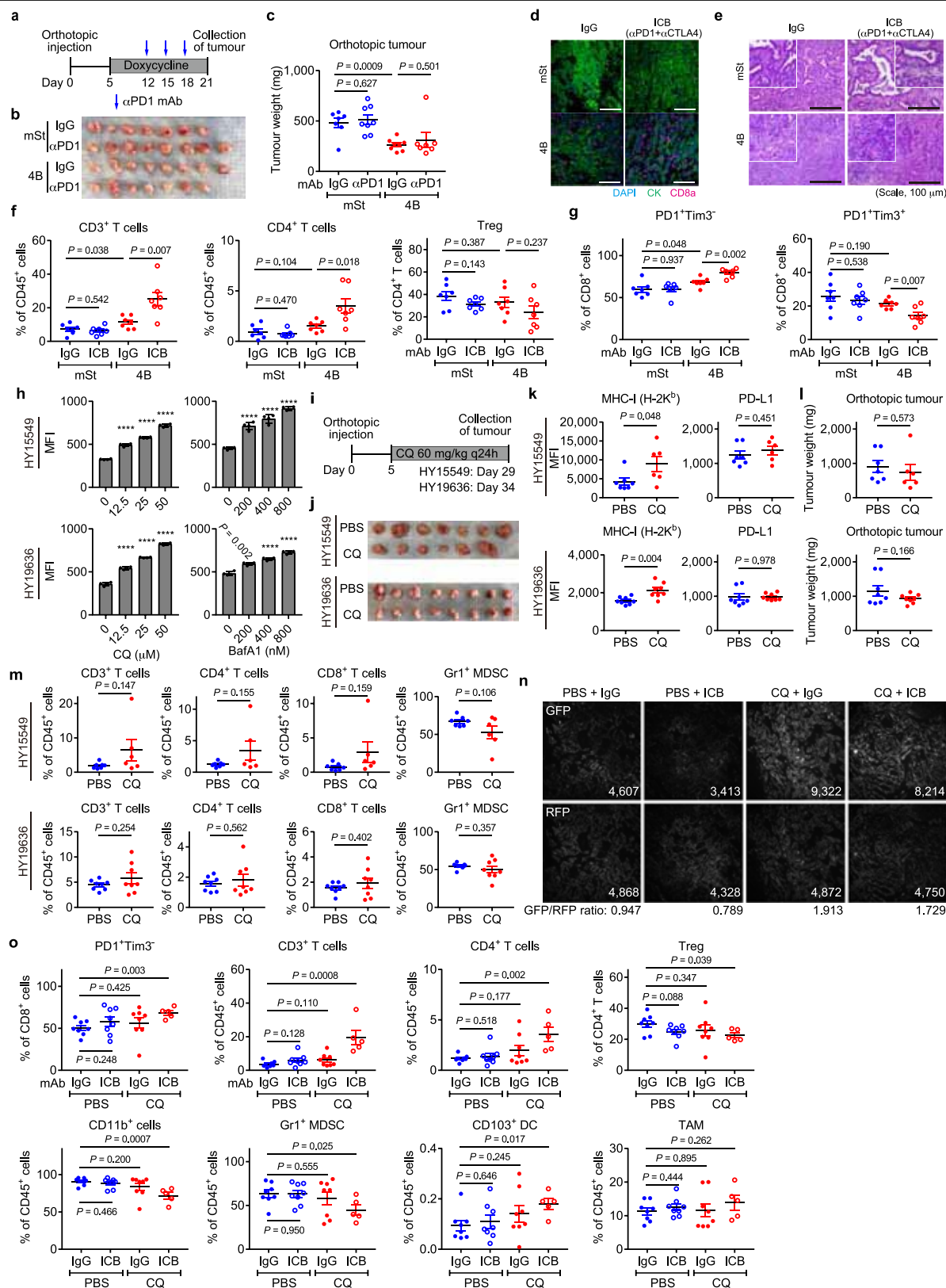
**Extended Data Fig. 8 | Separation of PDAC cells with distinct autophagy flux using the GFP-LC3-RFP reporter.** Heterogeneity in basal autophagy flux was explored using mouse PDAC cells (HY15549) expressing the GFP-LC3-RFP reporter. **a, b**, HY15549 cells were grown as organoids or transplanted into C57BL/6 mice to form orthotopic tumours. **a**, Autophagy flux, as represented by GFP/RFP ratio, was measured by flow cytometry. Atg5<sup>-/-</sup> mouse embryonic fibroblasts (MEF) with the GFP-LC3-RFP reporter (black) was used as a control. Representative flow plots from three independent experiments are shown. **b**, Representative fluorescent images of orthotopic tumours. Cells with high autophagy flux show GFP-LC3 puncta formation (inset, arrowhead) and a decrease in total GFP-fluorescent signals, displaying red appearance in the merged image. Scale bars, 100  $\mu$ m. **c-h**, Mouse PDAC organoids were dissociated into single cells and sorted into autophagy-high (ATHi) or -low (ATLo) cells according to the GFP/RFP ratio. **c**, Sorting strategies. **d**, Top KEGG

pathways enriched in ATHi cells compared to ATLo cells. Gene set enrichment analysis (GSEA) was performed using RNA sequencing (RNA-seq) data from sorted ATHi and ATLo cells ( $n = 2$  and 3 biologically independent samples), showing enrichment of the autophagy-lysosome gene signatures in ATHi cells as compared with the ATLo cells. FDR, false discovery rate; NES, normalized enrichment score; Nom., nominal. **e, f**, Relative mRNA expression of autophagy/lysosome-related genes in the respective populations sorted from pooled populations (**e**) or a single-cell derived clone (**f**).  $n = 3$  technical replicates. Representative results from four (**e**) and two (**f**) independent sorting experiments are shown. **g-i**, Clonogenic potential of sorted ATHi and ATLo cells (**g, h**) and PDAC cells with Dox-inducible ATG4B(C74A) (AY6284) (**i**). Representative data from at least two independent experiments are shown.  $n = 4$  (**g**) and  $n = 3$  (**i**) per group. Data are mean  $\pm$  s.d. (**e-i**).  $P$  values were determined by unpaired two-tailed  $t$ -tests.



**Extended Data Fig. 9 | Basal autophagy flux determines immunogenicity of PDAC cells.** Mouse PDAC cells (HY15549) expressing the GFP-LC3-RFP reporter were sorted into ATThi and ATLo cells (Extended Data Fig. 8c) and injected into the pancreas (**a–g**) or the spleen (**h–k**) of C57BL/6 mice (**a–f, h–k**) or nude mice (**g**). Cells were sorted from pooled populations except for (**b**). **a, b**, Tumour weight on day 21. ATThi and ATLo cells were sorted from either pooled populations (**a**) ( $n = 9$  and  $10$ ) or a single-cell derived clone (**b**) ( $n = 10$  per group). **c–e**, Tumours shown in **a** were analysed. **c**, Cell surface MHC-I levels on PDAC cells measured by flow cytometry. **d**, Correlation between MHC-I levels on PDAC cells and tumour weight. **e**, Representative images of H&E (left) and immunofluorescent staining (right). Scale bars,  $100 \mu\text{m}$ . **f**, Quantification of tumour-infiltrating immune cells by flow cytometry ( $n = 8$  per group). Orthotopic tumours obtained on day 21 were analysed. **g**, Orthotopic tumours

in nude mice obtained on day 19 ( $n = 8$  and  $7$ ). **h–k**, Liver metastasis model. **h**, Study design. **i, j**, Weight of livers (**i**) and cell surface MHC-I levels on PDAC cells measured by flow cytometry (**j**) on day 17 ( $n = 5$  per group). **k**, Representative immunofluorescence images of livers obtained on day 9 ( $n = 3$  per group). Frozen sections were stained with anti-CD8a antibody and DAPI. In these merged images, cells with high autophagy flux appear as red, reflecting the relative loss of GFP-fluorescence and lower GFP/RFP ratio, whereas cells with low autophagy flux appear as yellow to green, reflecting high GFP/RFP ratio. In the enlarged images, CD8a<sup>+</sup> cells were indicated by white arrowheads. Scale bars,  $100 \mu\text{m}$ . For **a, c–f** and **i–k**, experiments were performed at least twice and representative data of one experiment are shown. Data are mean  $\pm$  s.e.m. (**a–c, f, g, i, j**).  $n$  indicates individual mice.  $P$  values were determined by unpaired two-tailed  $t$ -tests (**a–c, f, g, i, j**) and Pearson correlation analysis (**d**).



**Extended Data Fig. 10** | See next page for caption.

**Extended Data Fig. 10 | Autophagy inhibition synergizes with dual ICB. a–c,** Anti-PD1 antibody treatment did not affect tumour growth in either control or autophagy-inhibited tumours. Mice bearing orthotopic PDAC tumours (HY15549) carrying Dox-inducible mSt or 4B were treated with Dox beginning on day 5 and received either isotype control IgG or anti-PD1 antibody ( $n = 7, 8, 8$  and 7 per group). **a**, Study design. **b, c**, Images (**b**) and weight (**c**) of tumours. **d–g**, Related to Fig. 4a–d. Mice bearing orthotopic PDAC tumours (HY15549) expressing Dox-inducible mSt or 4B were treated with Dox beginning on day 5 and received either isotype control IgG or dual ICB (anti-PD1/CTLA4 antibodies) ( $n = 7$  per group). **d, e**, Representative images of immunofluorescence staining (**d**) and H&E staining (**e**). Scale bars, 100  $\mu\text{m}$ . **f, g**, Quantification of tumour-infiltrating immune cells by flow cytometry. **h**, Cell surface MHC-I (H-2K<sup>b</sup>/D<sup>b</sup>) levels measured by flow cytometry. Mouse PDAC cells were treated with chloroquine or BafA1 at the indicated concentrations for 48 h ( $n = 4$ ). Mouse PDAC cells were grown in 2D culture (chloroquine) or as

organoids (BafA1). Representative results from at least three independent experiments are shown. **i–m**, Mice bearing orthotopic PDAC tumours expressing the GFP–LC3–RFP reporter were treated with PBS or chloroquine beginning on day 5 ( $n = 7$  vs 6 for HY15549 and  $n = 8$  vs 8 for HY19636). **i**, Study design. **j**, Images of tumours. **k**, Cell-surface MHC-I and PD-L1 levels on PDAC cells measured by flow cytometry. **l**, Tumour weight. **m**, Quantification of tumour-infiltrating immune cells by flow cytometry. **n**, Representative fluorescence images of tumours expressing the GFP–LC3–RFP reporter from Fig. 4j. Numerical values represent mean fluorescent intensity of each field. **o**, Quantification of tumour-infiltrating immune cells by flow cytometry ( $n = 8, 8, 8$  and 5; left to right). Tumours in Fig. 4k were analysed. Data are mean  $\pm$  s.e.m. (**c, f, g, k–m, o**) or  $\pm$  s.d. (**h**).  $n$  indicates individual mice (**c, f, g, k–m, o**) or biological replicates (**h**). Except for the orthotopic implantation of HY19636 cells (**j–m**), all experiments were performed at least twice and representative data of one experiment are shown. \*\*\*\* $P < 0.0001$ , unpaired two-tailed  $t$ -tests.



## Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

### Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

- |                                     |  |
|-------------------------------------|--|
| n/a                                 | Confirmed  |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement  |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly  |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> The statistical test(s) used AND whether they are one- or two-sided<br><i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i>   |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> A description of all covariates tested   |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons  |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> For null hypothesis testing, the test statistic (e.g. $F$ , $t$ , $r$ ) with confidence intervals, effect sizes, degrees of freedom and $P$ value noted<br><i>Give <math>P</math> values as exact values whenever suitable.</i>                            |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings  |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes  |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Estimates of effect sizes (e.g. Cohen's $d$ , Pearson's $r$ ), indicating how they were calculated  |

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

### Software and code

Policy information about [availability of computer code](#)

Data collection	Flow cytometry data were collected using FACSDiva (BD Biosciences, version 8.0.1). Quantitative PCR data was obtained with the Bio-Rad CFX96 Real-Time System and Bio-Rad CFX Manager (BIO-RAD, version 3.1.1517.0823). Fluorescent images were obtained using Leica Application Suite X software (Leica, version 2.0.0.14332.2). Immunoblotting images were obtained using ChemiDoc Touch Imaging System (BIO-RAD) and Image Lab Touch Software (BIO-RAD, version 2.3.0.07).
Data analysis	Flow cytometry data were analysed using FlowJo (FlowJo, version 10.4). Fluorescent images were analyzed using FIJI/ImageJ (version 1.52p) or Leica Application Suite X software (Leica, version 2.0.0.14332.2). Gene Set Enrichment Analysis was conducted using GSEA software (version 3.0) with gene sets derived from the KEGG pathway database in the MSigDB collections. Statistical analysis was performed using Graphpad Prism (GraphPad Software, version 7). For proteomics analysis, raw files were converted to mzXML using Raw File Reader (Thermo Fisher Scientific, version 3.0.77) and spectra were searched using the SEQUEST software package. Search results were filtered using the Linear Discriminant Analysis (LDA) function in MASS Package in R as described in Huttlin et al. Cell 143, 1174-1189 (2010).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

### Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

RNA-sequencing data are deposited to the Gene Expression Omnibus (GEO) data repository with accession number GSE145766. Source Data are provided for all

experiments. Other data that support the findings of this study are available on request from the corresponding author upon reasonable request.

## Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

## Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No sample size calculation was done either for in vitro or in vivo studies. For in vivo studies, sample sizes were determined based on our preliminary experiments. In our experience, n = 7-10 mice per group is sufficient to detect meaningful biological differences with good reproducibility.
Data exclusions	No data or animals were excluded from analysis.
Replication	Experiments were repeated multiple times with similar results as indicated in the figure legends. All critical in vivo experiments were reproduced using the same system (e.g. repeat of autophagy inhibition in the same cell line multiple times using the same in vivo assays) and/or using orthogonal approaches (e.g. repeated in multiple cell lines – HY19636, another C57BL/6-background mouse PDAC cell line; repeated in independent tumor models - orthotopic vs metastasis model; repeated with alternative approaches to decrease autophagy – Atg4b DN, shRNAs to Atg7, autophagy high/low populations, pharmacological - CQ; repeated using independent methods of impairing CD8 T-cells – CD8 depleting antibodies, immunodeficient nude mice, Batf3 KO mice). Additionally, we reproduced the critical autophagy inhibition experiment with MHC-I upregulation and T-cell infiltration in a second mouse facility.
Randomization	Prior to treatment initiation, mice bearing tumours were randomized. Tumour sizes were not measured at the time of randomization.
Blinding	Blinding was not performed in mouse experiments. The investigator needed to know the treatment groups in order to perform the study. Tumour weights (an objective measurement) were carried out only at the study endpoints after mice were euthanized and tumours were harvested.

## Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	Briefly describe the study type including whether data are quantitative, qualitative, or mixed-methods (e.g. qualitative cross-sectional, quantitative experimental, mixed-methods case study).
Research sample	State the research sample (e.g. Harvard university undergraduates, villagers in rural India) and provide relevant demographic information (e.g. age, sex) and indicate whether the sample is representative. Provide a rationale for the study sample chosen. For studies involving existing datasets, please describe the dataset and source.
Sampling strategy	Describe the sampling procedure (e.g. random, snowball, stratified, convenience). Describe the statistical methods that were used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient. For qualitative data, please indicate whether data saturation was considered, and what criteria were used to decide that no further sampling was needed.
Data collection	Provide details about the data collection procedure, including the instruments or devices used to record the data (e.g. pen and paper, computer, eye tracker, video or audio equipment) whether anyone was present besides the participant(s) and the researcher, and whether the researcher was blind to experimental condition and/or the study hypothesis during data collection.
Timing	Indicate the start and stop dates of data collection. If there is a gap between collection periods, state the dates for each sample cohort.
Data exclusions	If no data were excluded from the analyses, state so OR if data were excluded, provide the exact number of exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.
Non-participation	State how many participants dropped out/declined participation and the reason(s) given OR provide response rate OR state that no participants dropped out/declined participation.
Randomization	If participants were not allocated into experimental groups, state so OR describe how participants were allocated to groups, and if allocation was not random, describe how covariates were controlled.

# Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	Briefly describe the study. For quantitative data include treatment factors and interactions, design structure (e.g. factorial, nested, hierarchical), nature and number of experimental units and replicates.
Research sample	Describe the research sample (e.g. a group of tagged <i>Passer domesticus</i> , all <i>Stenocereus thurberi</i> within Organ Pipe Cactus National Monument), and provide a rationale for the sample choice. When relevant, describe the organism taxa, source, sex, age range and any manipulations. State what population the sample is meant to represent when applicable. For studies involving existing datasets, describe the data and its source.
Sampling strategy	Note the sampling procedure. Describe the statistical methods that were used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient.
Data collection	Describe the data collection procedure, including who recorded the data and how.
Timing and spatial scale	Indicate the start and stop dates of data collection, noting the frequency and periodicity of sampling and providing a rationale for these choices. If there is a gap between collection periods, state the dates for each sample cohort. Specify the spatial scale from which the data are taken
Data exclusions	If no data were excluded from the analyses, state so OR if data were excluded, describe the exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.
Reproducibility	Describe the measures taken to verify the reproducibility of experimental findings. For each experiment, note whether any attempts to repeat the experiment failed OR state that all attempts to repeat the experiment were successful.
Randomization	Describe how samples/organisms/participants were allocated into groups. If allocation was not random, describe how covariates were controlled. If this is not relevant to your study, explain why.
Blinding	Describe the extent of blinding used during data acquisition and analysis. If blinding was not possible, describe why OR explain why blinding was not relevant to your study.
Did the study involve field work?	<input type="checkbox"/> Yes <input type="checkbox"/> No

## Field work, collection and transport

Field conditions	Describe the study conditions for field work, providing relevant parameters (e.g. temperature, rainfall).
Location	State the location of the sampling or experiment, providing relevant parameters (e.g. latitude and longitude, elevation, water depth).
Access and import/export	Describe the efforts you have made to access habitats and to collect and import/export your samples in a responsible manner and in compliance with local, national and international laws, noting any permits that were obtained (give the name of the issuing authority, the date of issue, and any identifying information).
Disturbance	Describe any disturbance caused by the study and how it was minimized.

## Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

### Materials & experimental systems

n/a	Involved in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies
<input type="checkbox"/>	<input checked="" type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology
<input type="checkbox"/>	<input checked="" type="checkbox"/> Animals and other organisms
<input type="checkbox"/>	<input checked="" type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data

### Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input type="checkbox"/>	<input checked="" type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

All antibodies, their supplier, and the dilution used is reported in detail in the methods section and Supplementary Table 1.

Immunofluorescence (IF) (Target molecule, target species, host species, clone, conjugation, dilution, company, catalog number, Lot number)

LAMP1 Human Rabbit D2D11 unconjugated 1:500 Cell Signaling Technology 9091 Lot 6  
HLA-A, B, C Human Mouse W6/32 unconjugated 1:250 Sigma-Aldrich H1650 098M4897V  
LC3B Human Rabbit D11 unconjugated 1:500 Cell Signaling Technology 3868 Lot 13  
IgG secondary Human Rabbit N/A AF555 1:400 ThermoFisher Scientific A-21429 1937155  
IgG secondary Human Mouse N/A AF488 1:250 ThermoFisher Scientific A-11029 TJ269844

Immunoblotting (IB) (Target molecule, target species, host species, clone, dilution, company, catalog number, Lot number)

HLA-A, B, C Human Mouse EMR8-5 1:2000 Abcam ab70328 GR3266685-1  
HLA-A, B Human Rabbit EPR1394Y 1:1000 Abcam ab134189 GR119003-11  
HLA-A Human Rabbit EP1395Y 1:2000 Abcam ab52922 GR258732-23  
HLA-B Human Rabbit polyclonal 1:1500 ThermoFisher Scientific PA5-35345 UB2717115  
HLA-C Human Rabbit EPR6749 1:2500 Abcam ab126722 GR97813-8  
Niemann Pick C1 Human Rabbit EPR5209 1:3000 Abcam ab134113 GR3230423-7  
GAPDH Human Mouse 6C5 1:1000 Santa Cruz sc-32233 E2419  
Alpha Tubulin Human Mouse 1E4C11 1:5000 Proteintech 66031-1-Ig Lot# not available  
LAMP1 Human Mouse D2D11 1:1000 Cell Signaling Technology 9091 Lot 6  
ATG3 Human Rabbit polyclonal 1:1000 Cell Signaling Technology 3415 Lot 2  
ATG7 Human Rabbit D12B11 1:1000 Cell Signaling Technology 8558 Lot 3  
RB1CC1 (FIP200) Human Rabbit polyclonal 1:750 Protein-Tech 17250-1-AP Lot# not available  
ULK1 Mouse Rabbit D8H5 1:1000 Cell Signaling Technology 8054 Lot 6  
Rubicon Mouse Rabbit D9F7 1:1000 Cell Signaling Technology 8465 Lot 2  
ATG4B Mouse Rabbit polyclonal 1:1000 Cell Signaling Technology 5299 Lot 2  
NBR1 Human Mouse 4BR 1:500 Santa Cruz sc-130380 L0219  
Optineurin Human Rabbit polyclonal 1:500 Abcam ab23666 GR271296-15  
TAX1BP1 Human Rabbit polyclonal 1:1000 Sigma-Aldrich HPA024432 B118261  
CALCOCO2/NDP52 Human Rabbit polyclonal 1:1000 Sigma-Aldrich HPA023195 A65497  
p62/SQSTM1 Human Guinea Pig polyclonal 1:1000 Progen GP62-C DS-190318  
ATG7 Mouse Rabbit polyclonal 1:1000 Sigma-Aldrich A2856 076M4824V  
LC3B Mouse Rabbit polyclonal 1:1000 Novus Biologicals NB100-220 EM  
RFP/mStrawberry Rabbit polyclonal 1:1000 Rockland 600-401-379 35055  
β-actin Mouse Mouse AC-15 1:1000 Sigma-Aldrich A5441 127M4866V  
Mono- and polyubiquitinated conjugates Human Mouse FK2 1:1000 Enzo BML-PW8810-0100 1031904

Immunohistochemistry (IHC) (Target molecule, target species, host species, clone, conjugation, dilution, company, catalog number, Lot number)

CK-19 Human Rat TROMA-III unconjugated 1:300 DSHB Lot# not available  
HLA-A,B Human Rabbit EPR1394Y unconjugated 1:300 abcam ab134189 GR119003-11  
pan-Cytokeratin Mouse Mouse C-11 AF488 1:200 BioLegend 628608 B264191  
CD8a Mouse Rat 4SM15 unconjugated 1:100 Fisher 14-0808-80 2144532  
RFP/mStrawberry N/A Rabbit polyclonal unconjugated 1:400 Rockland 600-401-379 35055  
CD3 Mouse Rat 17A2 unconjugated 1:100 BioLegend 100201 B229443  
CD8a Mouse Rat 53-6.7 unconjugated 1:100 BioLegend 100701 B267129  
CD8a Mouse Rat 53-6.7 AF647 1:100 BioLegend 100727 B262298  
Rat IgG Rat Goat polyclonal AF647 1:400 Thermo Fisher A-21247 2119156  
Rabbit IgG Rabbit Goat polyclonal AF555 1:400 Thermo Fisher A-21428 1937155

Flow Cytometry (FCM) (Target molecule, target species, host species, clone, conjugation, dilution, company, catalog number, Lot number)

HLA-A, B, C Human Mouse W6/32 AF488 1:75 BioLegend 311413 B256489  
HLA-A, B, C Human Mouse W6/32 PE 1:75 BioLegend 311406 B265864  
CD4 Mouse Rat GK1.5 BV605 1:100 BioLegend 100451 B258998  
CD45 Mouse Rat 30-F11 BUV395 1:100 BD Biosciences 564279 9171873  
Foxp3 Mouse Rat FJK-16s eFluor 450 1:25 eBioscience 48-5773-82 1988692  
CD3 Mouse Rat 17A2 BV711 1:100 BioLegend 100241 B265870  
B220 Mouse Rat RA3-6B2 APC 1:100 BioLegend 103212 B242829  
Nk1.1 Mouse Mouse PK136 APC/Cy7 1:100 BioLegend 108723 B261757  
CD11b Mouse Rat M1/70 PE/Cy7 1:100 BioLegend 101215 B268266  
CD11c Mouse Armenian Hamster N418 Pacific 1:100 BioLegend 117321 B257457  
CD103 Mouse Armenian Hamster 2E7 AF488 1:100 BioLegend 121407 B276138  
F4/80 Mouse Rat BM8 BV605 1:100 BioLegend 123133 B268952  
Gr1 Mouse Rat RB6-8C5 APC/Cy7 1:100 BioLegend 108423 B271583  
CD8a Mouse Rat 53-6.7 AF700 1:100 BioLegend 100730 B246153  
CD8a Mouse Rat 53-6.7 AF647 1:200 BioLegend 100727 B262298  
MHC-II Mouse Rat M5/114.15.2 AF700 1:100 BioLegend 107622 B264454  
PD-1 Mouse Rat 29F.1A12 BV785 1:100 BioLegend 135225 B261641  
Tim3 Mouse Rat RMT3-23 APC 1:100 BioLegend 119705 B265068  
PD-L1 Mouse Rat 10F.9G2 BV711 1:100 BioLegend 124319 B257887

H-2 Mouse Rat M1/42 FITC 1:100 BioLegend 125507 B247755  
H-2Kb Mouse Mouse AF6-88.5 AF647 1:100 BioLegend 116512 B265968  
H-2Kb/H-2Db Mouse Mouse 28-8-6 AF647 1:100 BioLegend 114612 B253985  
H-2Kb bound to SIINFEKL Mouse Mouse 25-D1.16 APC 1:100 BioLegend 141605 B253603  
H-2Db Mouse Mouse KH95 APC 1:100 BioLegend 111513 B269797  
Transferrin receptor Mouse Rat RI7217 APC 1:100 BioLegend 113819 B264912  
HVEM Mouse Armenian Hamster HMHV-1B18 APC 1:100 BioLegend 136305 B266707  
Galectin-3 Mouse Rat M3/38 APC 1:100 BioLegend 125419 B272403  
Galectin-9 Mouse Rat RG9-35 APC 1:100 BioLegend 136109 B245181  
E-Cadherin Mouse Rat DECMA-1 AF647 1:100 BioLegend 147307 B220119  
CD44 Mouse Rat IM7 APC 1:100 BioLegend 103012 B265921  
TNFRSF1A Mouse Armenian Hamster 55R-286 APC 1:100 BioLegend 113005 B255749  
EpCAM Mouse Rat G8.8 APC 1:100 BioLegend 118213 B249071  
Fas (CD95) Mouse Mouse SA367H8 APC 1:100 BioLegend 152603 B278865  
DR5 Mouse Armenian Hamster MD5-1 PE 1:100 BioLegend 119905 B260464  
CD155 Mouse Rat TX56 APC 1:100 BioLegend 131509 B284616  
PD-L1 Mouse Rat 10F.9G2 APC 1:100 BioLegend 124311 B241508

Blocking antibodies for in vitro and in vivo experiments (name, host species, clone, dilution, company, catalog number, Lot number).

InVivoMab anti-mouse CD8a Rat 53-6.7 200 mg/mouse BioXCell BE0004-1 693018M2  
InVivoMab rat IgG2a isotype control Rat 2A3 200 mg/mouse BioXCell BE0089 716719J3  
InVivoMab anti-mouse PD-1 Rat RMP1-14 200 mg/mouse BioXCell BE0146 735019J3  
InVivoMab polyclonal Syrian hamster IgG Hamster N/A 200 mg/mouse BioXCell BE0087 709119A1  
InVivoMab anti-mouse CTLA-4 Hamster 9H10 200 mg/mouse BioXCell BE0131 704019A2  
InVivoMab anti-mouse MHC Class I (H-2Kb) bound to SIINFEKL peptide Rat 25-D1.16 BioXCell BE0207 507816J3  
InVivoMab mouse IgG1 isotype control Mouse MOPC-21 BioXCell BE0083 701618J2

## Validation

All antibodies used are commercially available and validated by the manufacturers. For the proteins listed below, specificities of the antibodies were validated using appropriate controls (cells deficient for or overexpressing the target molecules, cells treated with reagents known to increase or decrease the protein, or isolated organelles with specific protein expression, etc.): LAMP1, LC3B, ATG3, ATG7, RB1CC1 (FIP200), ULK1, Rubicon, ATG4B, NBR1, RFP, mStrawberry, mono- and polyubiquitinated proteins, H-2, H-2Kb, H-2Db, H-2Kb/Db, and H-2Kb bound to SIINFEKL. For flow cytometry, fluorescence minus one (FMO) controls and isotype controls were used to accurately determine the thresholds.

Additional information can be found on the manufactures' websites: (Target Antigen, Company, Catalog #, URL)

### Immunofluorescence (IF)

LAMP1 Cell Signaling Technology 9091 <https://www.cellsignal.com/products/primary-antibodies/lamp1-d2d11-xp-rabbit-mab/9091>  
HLA-A, B, C Sigma-Aldrich H1650 <https://www.sigmaaldrich.com/catalog/product/sigma/h1650>  
LC3B Cell Signaling Technology 3868 <https://www.cellsignal.com/products/primary-antibodies/lc3b-d11-xp-rabbit-mab/3868>

### Immunoblotting (IB)

HLA-A, B, C Abcam ab70328 <https://www.abcam.com/hla-class-1-abc-antibody-emr8-5-ab70328.html>  
HLA-A, B Abcam ab134189 <https://www.abcam.com/mhc-class-i--hla-a--hla-b-antibody-epr1394y-ab134189.html>  
HLA-A Abcam ab52922 <https://www.abcam.com/hla-a-antibody-ep1395y-ab52922.html>  
HLA-B ThermoFisher Scientific PA5-35345 <https://www.thermofisher.com/antibody/product/HLA-B-Antibody-Polyclonal/PA5-35345>  
HLA-C Abcam ab126722 <https://www.abcam.com/hla-c-antibody-epr6749-ab126722.html>  
Niemann Pick C1 Abcam ab134113 <https://www.abcam.com/niemann-pick-c1-antibody-epr5209-ab134113.html>  
GAPDH Santa Cruz sc-32233 <https://www.scbt.com/p/gapdh-antibody-6c5;jsessionid=O9eiHtvkm7s7v2yLUzMlb8KDQ3CExutf-IM9rXpJ8JoNEVczZjRl-1044019884!2132931255>  
Alpha Tubulin Proteintech 66031-1-Ig <https://www.ptglab.com/products/tubulin-Alpha-Antibody-66031-1-Ig.htm>  
LAMP1 Cell Signaling Technology 9091 <https://www.cellsignal.com/products/primary-antibodies/lamp1-d2d11-xp-rabbit-mab/9091>  
ATG3 Cell Signaling Technology 3415 <https://www.cellsignal.com/products/primary-antibodies/atg3-antibody/3415>  
ATG7 Cell Signaling Technology 8558 <https://www.cellsignal.com/products/primary-antibodies/atg7-d12b11-rabbit-mab/8558>  
RB1CC1 (FIP200) Protein-Tech 17250-1-AP <https://www.ptglab.com/Products/RB1CC1-Antibody-17250-1-AP.htm>  
ULK1 Cell Signaling Technology 8054 <https://www.cellsignal.com/products/primary-antibodies/ulk1-d8h5-rabbit-mab/8054>  
Rubicon Cell Signaling Technology 8465 <https://www.cellsignal.com/products/primary-antibodies/rubicon-d9f7-rabbit-mab/8465>  
ATG4B Cell Signaling Technology 5299 <https://www.cellsignal.com/products/primary-antibodies/atg4b-antibody/5299>  
NBR1 Santa Cruz sc-130380 [https://www.scbt.com/p/nbr1-antibody-4br?productCanUrl=nbr1-antibody-4br&\\_requestid=2564925](https://www.scbt.com/p/nbr1-antibody-4br?productCanUrl=nbr1-antibody-4br&_requestid=2564925)  
Optineurin Abcam ab23666 <https://www.abcam.com/optineurin-antibody-ab23666.html>  
TAX1BP1 Sigma-Aldrich HPA024432 <https://www.sigmaaldrich.com/catalog/product/sigma/hpa024432>  
CALCOCO2/NDP52 Sigma-Aldrich HPA023195 <https://www.sigmaaldrich.com/catalog/product/sigma/hpa023195>  
p62/SQSTM1 Progen GP62-C <https://www.progen.com/anti-p62-sqstm1-c-terminus-guinea-pig-polyclonal-serum.html>  
ATG7 Sigma-Aldrich A2856 <https://www.sigmaaldrich.com/catalog/product/sigma/a2856>  
LC3B Novus Biologicals NB100-220 [https://www.novusbio.com/products/lc3b-antibody\\_nb100-2220](https://www.novusbio.com/products/lc3b-antibody_nb100-2220)  
RFP/mStrawberry Rockland 600-401-379 [https://rockland-inc.com/store/Antibodies-to-GFP-and-Antibodies-to-RFP-600-401-379-O4L\\_24299.aspx](https://rockland-inc.com/store/Antibodies-to-GFP-and-Antibodies-to-RFP-600-401-379-O4L_24299.aspx)  
β-actin Sigma-Aldrich A5441 <https://www.sigmaaldrich.com/catalog/product/sigma/a5441>  
Mono- and polyubiquitinated conjugates Enzo BML-PW8810-0100 <https://www.enzolifesciences.com/BML-PW8810/mono-and-polyubiquitinated-conjugates-mono-clonal-antibody-fk2/>



Immunohistochemistry (IHC)  
 CK-19 DSHB TROMA-III <https://dshb.biology.uiowa.edu/TROMA-III>  
 HLA-A,B abcam ab134189 <https://www.abcam.com/mhc-class-i--hla-a--hla-b-antibody-epr1394y-ab134189.html>  
 pan-Cytokeratin BioLegend 628608 <https://www.biolegend.com/en-us/products/alexa-fluor-488-anti-cytokeratin-pan-reactive-antibody-13383>  
 CD8a Fisher 14-0808-80 <https://www.thermofisher.com/antibody/product/CD8a-Antibody-clone-4SM15-Monoclonal/14-0808-82>  
 RFP/mStrawberry Rockland 600-401-379 [https://rockland-inc.com/store/Antibodies-to-GFP-and-Antibodies-to-RFP-600-401-379-O4L\\_24299.aspx](https://rockland-inc.com/store/Antibodies-to-GFP-and-Antibodies-to-RFP-600-401-379-O4L_24299.aspx)  
 CD3 BioLegend 100201 <https://www.biolegend.com/en-us/products/purified-anti-mouse-cd3-antibody-48>  
 CD8a BioLegend 100701 <https://www.biolegend.com/en-us/products/purified-anti-mouse-cd8a-antibody-157>  
 CD8a BioLegend 100727 <https://www.biolegend.com/en-us/products/alexa-fluor-647-anti-mouse-cd8a-antibody-2699>

Flow Cytometry (FCM)  
 HLA-A, B, C BioLegend 311413 <https://www.biolegend.com/en-us/products/alexa-fluor-488-anti-human-hla-a-b-c-antibody-2899>  
 HLA-A, B, C BioLegend 311406 <https://www.biolegend.com/en-us/products/pe-anti-human-hla-a-b-c-antibody-1872>  
 CD4 BioLegend 100451 <https://www.biolegend.com/en-us/products/brilliant-violet-605-anti-mouse-cd4-antibody-10708>  
 CD45 BD Biosciences 564279 <https://www.bdbiosciences.com/us/reagents/research/antibodies-buffers/immunology-reagents/anti-mouse-antibodies/cell-surface-antigens/buv395-rat-anti-mouse-cd45-30-f11/p/564279>  
 Foxp3 eBioscience 48-5773-82 <https://www.thermofisher.com/antibody/product/FOXP3-Antibody-clone-FJK-16s-Monoclonal/48-5773-82>  
 CD3 BioLegend 100241 <https://www.biolegend.com/en-us/products/brilliant-violet-711-anti-mouse-cd3-antibody-10022>  
 B220 BioLegend 103212 <https://www.biolegend.com/en-us/products/apc-anti-mouse-human-cd45r-b220-antibody-442>  
 Nk1.1 BioLegend 108723 <https://www.biolegend.com/en-us/products/apc-cy7-anti-mouse-nk-1-1-antibody-4002>  
 CD11b BioLegend 101215 <https://www.biolegend.com/en-us/products/apc-cy7-anti-mouse-nk-1-1-antibody-4002>  
 CD11c BioLegend 117321 <https://www.biolegend.com/en-us/products/pacific-blue-anti-mouse-cd11c-antibody-3864>  
 CD103 BioLegend 121407 <https://www.biolegend.com/en-us/products/alexa-fluor-488-anti-mouse-cd103-antibody-3575>  
 F4/80 BioLegend 123133 <https://www.biolegend.com/en-us/products/brilliant-violet-605-anti-mouse-f4-80-antibody-8702>  
 Gr1 BioLegend 108423 <https://www.biolegend.com/en-us/products/apccyanine7-anti-mouse-ly-6gry-6c-gr-1-antibody-3935>  
 CD8a BioLegend 100730 <https://www.biolegend.com/en-us/products/alexa-fluor-700-anti-mouse-cd8a-antibody-3387>  
 CD8a BioLegend 100727 <https://www.biolegend.com/en-us/products/alexa-fluor-647-anti-mouse-cd8a-antibody-2699>  
 MHC-II BioLegend 107622 <https://www.biolegend.com/en-us/products/alexa-fluor-700-anti-mouse-i-a-i-e-antibody-3413>  
 PD-1 BioLegend 135225 <https://www.biolegend.com/en-us/products/brilliant-violet-785-anti-mouse-cd279-pd-1-antibody-9874>  
 Tim3 BioLegend 119705 <https://www.biolegend.com/en-us/products/apc-anti-mouse-cd366-tim-3-antibody-8238>  
 PD-L1 BioLegend 124319 <https://www.biolegend.com/en-us/products/brilliant-violet-711-anti-mouse-cd274-b7-h1--pd-l1-antibody-9808>  
 H-2 BioLegend 125507 <https://www.biolegend.com/en-us/products/fitc-anti-mouse-h-2-antibody-5346>  
 H-2Kb BioLegend 116512 <https://www.biolegend.com/en-us/products/alexa-fluor-647-anti-mouse-h-2kb-antibody-3238>  
 H-2Kb/H-2Db BioLegend 114612 <https://www.biolegend.com/en-us/products/alexa-fluor-647-anti-mouse-h-2kb-h-2db-antibody-3222>  
 H-2Kb bound to SIINFEKL BioLegend 141605 <https://www.biolegend.com/en-us/products/apc-anti-mouse-h-2kb-bound-to-siinfekl-antibody-7882>  
 H-2Db BioLegend 111513 <https://www.biolegend.com/en-us/products/apc-anti-mouse-h-2d-b-antibody-11947>  
 Transferrin receptor BioLegend 113819 <https://www.biolegend.com/en-us/products/apc-anti-mouse-cd71-antibody-15498>  
 HVEM BioLegend 136305 <https://www.biolegend.com/en-us/products/apc-anti-mouse-cd270-hvem-antibody-6419>  
 Galectin-3 BioLegend 125419 <https://www.biolegend.com/en-us/products/apc-anti-mouse-human-mac-2-galectin-3-antibody-15145>  
 Galectin-9 BioLegend 136109 <https://www.biolegend.com/en-us/products/apc-anti-mouse-galectin-9-antibody-8239>  
 E-Cadherin BioLegend 147307 <https://www.biolegend.com/en-us/products/alexa-fluor-647-anti-mouse-human-cd324-e-cadherin-antibody-10024>  
 CD44 BioLegend 103012 <https://www.biolegend.com/en-us/products/apc-anti-mouse-human-cd44-antibody-312>  
 TNFRSF1A (CD120a) BioLegend 113005 <https://www.biolegend.com/en-us/products/apc-anti-mouse-cd120a-tnf-r-type-i-p55-antibody-6268>  
 EpCAM BioLegend 118213 <https://www.biolegend.com/en-us/products/apc-anti-mouse-cd326-ep-cam-antibody-4974>  
 Fas (CD95) BioLegend 152603 <https://www.biolegend.com/en-us/products/apc-anti-mouse-cd95-fas-antibody-13906>  
 DR5 BioLegend 119905 <https://www.biolegend.com/en-us/products/pe-anti-mouse-cd262-dr5--trail-r2-antibody-2668>  
 CD155 BioLegend 131509 <https://www.biolegend.com/en-us/products/apc-anti-mouse-cd155-pvr-antibody-6722>  
 PD-L1 BioLegend 124311 <https://www.biolegend.com/en-us/products/apc-anti-mouse-cd274-b7-h1--pd-l1-antibody-6655>

## Eukaryotic cell lines

### Policy information about cell lines

#### Cell line source(s)

The cell lines PaTu-8988T, KP4, MiaPaca2, Panc 2.03, PaTu-8902, Panc1, AsPc1, HupT3, A549 were obtained from the American Type Culture Collection (ATCC) or the DSMZ. H441, H358, HCT116 and BEAS-2B were provided by Eric Collisson, UCSF. HPDE was provided by Ming Tsao (ref 42). Primary murine PDAC cell lines were established from pancreatic tumours in respective GEMMs. HY15549 and HY19636 were derived from female KPC mice (p48-Cre+;KrasLSL-G12D/+;Trp53lox/+) that were fully backcrossed into a C57BL/6 background. Atg5+/+ KPC cells were derived from Pdx1-Cre+; KrasLSL-G12D/+;Trp53lox/++;Atg5+/+ mice. Atg5-/- KPC cells were derived from Pdx1-Cre+;KrasLSL-G12D/+;Trp53lox/++;Atg5lox/lox mice (ref 10). AY6284 cells were derived from p48-Cre+;KrasLSL-G12D/+;Trp53lox/++;RosaLSL-rtTA;Mst-Atg4BC74A mice (ref 12).

#### Authentication

All human cell lines acquired from ATCC or DMSZ were validated by STR profiling, visual inspection, and careful maintenance in a central lab cell bank.

Mycoplasma contamination	Routine testing for mycoplasma was conducted by PCR. All cell lines tested negative for mycoplasma.
Commonly misidentified lines (See <a href="#">ICLAC</a> register)	No commonly misidentified cell lines were used.

## Palaeontology

Specimen provenance	<i>Provide provenance information for specimens and describe permits that were obtained for the work (including the name of the issuing authority, the date of issue, and any identifying information).</i>
Specimen deposition	<i>Indicate where the specimens have been deposited to permit free access by other researchers.</i>
Dating methods	<i>If new dates are provided, describe how they were obtained (e.g. collection, storage, sample pretreatment and measurement), where they were obtained (i.e. lab name), the calibration program and the protocol for quality assurance OR state that no new dates are provided.</i>

☐ Tick this box to confirm that the raw and calibrated dates are available in the paper or in Supplementary Information.

## Animals and other organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research

Laboratory animals	Female 8-10 week old C57BL/6 mice or NCr nude (B6NTac, Taconic) were purchased from Taconic and used for allograft experiments. OT-I transgenic mice (Cat# 003831) and Batf3-/- mice (Cat# 013755) were purchased from Jackson Laboratory.
Wild animals	This study did not involve wild animals.
Field-collected samples	This study did not involve samples collected in the field.
Ethics oversight	All experiments involving laboratory animals were approved by the New York University Institutional Animal Care and Use Committee (IACUC) under protocol numbers IA16-00507 and IA16-01331. This is indicated in the methods section.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

## Human research participants

Policy information about [studies involving human research participants](#)

Population characteristics	Human PDAC specimens were obtained from patients who underwent surgical resection of the primary PDAC tumours. Clinical characteristics of the patients are described in Supplementary Table 3.
Recruitment	Surgical resection of PDAC patients at UCSF were used for the study under an Institutional Review Board (IRB)-approved protocol (18-25787) at UCSF. The study had minimal risk per the IRB protocol and thus informed consent was not necessary.
Ethics oversight	Samples were collected at UCSF under IRB-approved protocol 18-25787. This study was conducted under strict compliance with institutional ethical regulations. The study had minimal risk per the IRB protocol and thus informed consent was not necessary.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

## Clinical data

Policy information about [clinical studies](#)

All manuscripts should comply with the ICMJE [guidelines for publication of clinical research](#) and a completed [CONSORT checklist](#) must be included with all submissions.

Clinical trial registration	n/a
Study protocol	n/a
Data collection	n/a
Outcomes	n/a

## ChIP-seq

### Data deposition

- ☐ Confirm that both raw and final processed data have been deposited in a public database such as [GEO](#).
- ☐ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

## Data access links

May remain private before publication.

For "Initial submission" or "Revised version" documents, provide reviewer access links. For your "Final submission" document, provide a link to the deposited data.

## Files in database submission

Provide a list of all files available in the database submission.

## Genome browser session

(e.g. [UCSC](#))

Provide a link to an anonymized genome browser session for "Initial submission" and "Revised version" documents only, to enable peer review. Write "no longer applicable" for "Final submission" documents.

## Methodology

## Replicates

Describe the experimental replicates, specifying number, type and replicate agreement.

## Sequencing depth

Describe the sequencing depth for each experiment, providing the total number of reads, uniquely mapped reads, length of reads and whether they were paired- or single-end.

## Antibodies

Describe the antibodies used for the ChIP-seq experiments; as applicable, provide supplier name, catalog number, clone name, and lot number.

## Peak calling parameters

Specify the command line program and parameters used for read mapping and peak calling, including the ChIP, control and index files used.

## Data quality

Describe the methods used to ensure data quality in full detail, including how many peaks are at FDR 5% and above 5-fold enrichment.

## Software

Describe the software used to collect and analyze the ChIP-seq data. For custom code that has been deposited into a community repository, provide accession details.

## Flow Cytometry

## Plots

Confirm that:

- ☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
- ☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).
- ☒ All plots are contour plots with outliers or pseudocolor plots.
- ☒ A numerical value for number of cells or percentage (with statistics) is provided.

## Methodology

## Sample preparation

For surface and intracellular MHC-I staining of human cell lines, cells were stained with Alexa Fluor® 488 anti-human HLA-A, B, C antibody (BioLegend, clone W6/32) at a 1:75 dilution for 45 minutes at 4°C in the dark and washed with PBS plus 2% FBS and 2mM EDTA (FACS buffer). Cells were then fixed and permeabilized prior to staining with PE conjugated anti-human HLA-A, B, C antibody (BioLegend, clone W6/32) at a 1:75 dilution for 45 min at room temperature and washed with FACS buffer. For cell surface molecule staining of mouse cells, single cell suspensions were washed with FCM buffer (HBSS containing 1% FBS, 1 mM EDTA, and 10 mM HEPES) and stained with antibodies (Supplementary Table 1) at 4°C in the dark for 20 min. Dead cells were depleted by DAPI staining.

For the immunophenotyping of tumours, tissues were mechanically minced with scissors, and then digested in DMEM containing 1 mg/mL Collagenase IV (Gibco), 100 µg/mL DNase I (Roche), 1% FBS, 10 mM HEPES, 2% Antibiotic-Antimycotic (ThermoFisher Scientific) for 40 min at 37°C in the dark with gentle agitation every 10 min. Digested tissues were then washed twice in DMEM containing 10% FBS, filtered through a 40-µm nylon mesh strainer (Corning). Cells were suspended in ACK Lysing Buffer (ThermoFisher Scientific), incubated for 10 min at 4°C in the dark to remove red blood cells. Cells were washed twice in FCM buffer and counted. Cells were stained with Zombie Aqua Fixable Viability Kit (BioLegend) and blocked with anti-mouse CD16/CD32 antibody (Mouse BD Fc Block™, clone 2.4G2, BD Bioscience). One million cells were incubated with appropriate antibodies (Supplementary Table 1) diluted in FCM buffer at 4°C in the dark for 40 min. Cells were then washed twice with FCM buffer and analyzed or further fixed in phosphate-buffered saline (PBS) containing 0.1 M L-Lysine (Sigma) and 2% paraformaldehyde (Ted Pella) at 4°C in the dark overnight. Intracellular staining was performed using Foxp3/Transcription Factor Staining Buffer Set (eBioscience).

## Instrument

Cells were analyzed on a BD LSR Fortessa or a BD LSR-II UV and analyzed by FlowJo software (FlowJo, LLC, version 10.4).

## Software

FACSDiva software (BD Biosciences) was used to collect the data, and the data was analyzed by FlowJo software (FlowJo, LLC, version 10.4).

## Cell population abundance

The purity of cells after sorting was confirmed by post-sort analyses and was usually &gt; 90%.

## Gating strategy

Gating strategies were described in Extended Data Fig. 7 and Supplementary Table 2.

- ☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.

# Magnetic resonance imaging

## Experimental design

Design type	<i>Indicate task or resting state; event-related or block design.</i>
Design specifications	<i>Specify the number of blocks, trials or experimental units per session and/or subject, and specify the length of each trial or block (if trials are blocked) and interval between trials.</i>
Behavioral performance measures	<i>State number and/or type of variables recorded (e.g. correct button press, response time) and what statistics were used to establish that the subjects were performing the task as expected (e.g. mean, range, and/or standard deviation across subjects).</i>

## Acquisition

Imaging type(s)	<i>Specify: functional, structural, diffusion, perfusion.</i>
Field strength	<i>Specify in Tesla</i>
Sequence & imaging parameters	<i>Specify the pulse sequence type (gradient echo, spin echo, etc.), imaging type (EPI, spiral, etc.), field of view, matrix size, slice thickness, orientation and TE/TR/flip angle.</i>
Area of acquisition	<i>State whether a whole brain scan was used OR define the area of acquisition, describing how the region was determined.</i>
Diffusion MRI	<input type="checkbox"/> Used <input type="checkbox"/> Not used

## Preprocessing

Preprocessing software	<i>Provide detail on software version and revision number and on specific parameters (model/functions, brain extraction, segmentation, smoothing kernel size, etc.).</i>
Normalization	<i>If data were normalized/standardized, describe the approach(es): specify linear or non-linear and define image types used for transformation OR indicate that data were not normalized and explain rationale for lack of normalization.</i>
Normalization template	<i>Describe the template used for normalization/transformation, specifying subject space or group standardized space (e.g. original Talairach, MNI305, ICBM152) OR indicate that the data were not normalized.</i>
Noise and artifact removal	<i>Describe your procedure(s) for artifact and structured noise removal, specifying motion parameters, tissue signals and physiological signals (heart rate, respiration).</i>
Volume censoring	<i>Define your software and/or method and criteria for volume censoring, and state the extent of such censoring.</i>

## Statistical modeling & inference

Model type and settings	<i>Specify type (mass univariate, multivariate, RSA, predictive, etc.) and describe essential details of the model at the first and second levels (e.g. fixed, random or mixed effects; drift or auto-correlation).</i>
Effect(s) tested	<i>Define precise effect in terms of the task or stimulus conditions instead of psychological concepts and indicate whether ANOVA or factorial designs were used.</i>
Specify type of analysis:	<input type="checkbox"/> Whole brain <input type="checkbox"/> ROI-based <input type="checkbox"/> Both
Statistic type for inference (See <a href="#">Eklund et al. 2016</a> )	<i>Specify voxel-wise or cluster-wise and report all relevant parameters for cluster-wise methods.</i>
Correction	<i>Describe the type of correction and how it is obtained for multiple comparisons (e.g. FWE, FDR, permutation or Monte Carlo).</i>

## Models & analysis

n/a	Involved in the study
<input type="checkbox"/>	<input type="checkbox"/> Functional and/or effective connectivity
<input type="checkbox"/>	<input type="checkbox"/> Graph analysis
<input type="checkbox"/>	<input type="checkbox"/> Multivariate modeling or predictive analysis
Functional and/or effective connectivity	<i>Report the measures of dependence used and the model details (e.g. Pearson correlation, partial correlation, mutual information).</i>
Graph analysis	<i>Report the dependent variable and connectivity measure, specifying weighted graph or binarized graph, subject- or group-level, and the global and/or node summaries used (e.g. clustering coefficient, efficiency, etc.).</i>







# KINDNESS UNDER CORONAVIRUS

COVID-19 lockdowns could lead to a more respectful research culture. **By Gemma Derrick**

**O**n 14 March, Spanish Prime Minister Pedro Sánchez announced a state of alarm (*Estado de alarma*) and a nationwide COVID-19 lockdown. This closed schools and led to some of Europe's most severe social-distancing rules, punishable by fines and enforced by a constant on-street presence of police and the army, often using drones. Similar restrictions had already been seen in China and Italy, and in the following weeks other European countries – including the United Kingdom – as well as the United States and my native Australia all followed suit.

Spain is where I found myself, a higher-education researcher, on lockdown day – weeks into a five-month research visit

to the country's national research council. I was away from home and my job at Lancaster University, UK, unable to go outside, and juggling the homeschooling of my five-year-old son in central Madrid for an unknown amount of time.

But alongside fear, I immediately witnessed an outpouring of kindness from colleagues and from the scientific community. My work focuses on building a kinder, gentler, more inclusive research culture by modifying one of its harshest processes, peer review. I know all too well how science can be an ultra-competitive, Machiavellian environment that can be unforgiving to individuals who are in need of guidance or generosity. However, in the light of COVID-19, both individual researchers and

research organizations are showing that there is a way of doing what we do that is not driven by ambition and ego.

Might a disease that forces us apart through social distancing make us closer in the long run? Here is a list of things I have experienced that make me think that kindness in research is possible:

- The online platforms I use for work (Microsoft Teams, Zoom, Skype and, more than ever, Twitter) are full of messages of support and unity from colleagues across the world, and with cheerful puppy pictures and virtual coffee breaks that are better attended than the physical variety.

- A once narrow, competitive drive to collect data for our individual research ambitions has been replaced by dialogue about whether now, with everyone's minds otherwise occupied, is really the best time to be collecting data? This goes beyond concerns about data reliability and reproducibility, towards a type of empathy and foresight that is the bedrock of research kindness.

- Deadlines for grant proposals have been relaxed and postponed. And normally overly bureaucratic procedures, such as lodging requests for no-cost extensions on grants to accommodate interrupted research schedules, have been met with promises that such extensions will be given, no questions asked.

- An e-mail from my university's ethics committee demanding changes to my application within a strict time frame was followed up, after my out-of-office message, with a "don't worry about your ethics application, stay safe", allowing a blessed moment of relief.

- As flexible-working requests have been granted and home working enforced, expectations about outputs have been replaced with empathy and understanding, an acknowledgement of caring responsibilities or sickness, and anxiety about our careers, families, friends and the world.

- On 24 March, the 2021 UK Research Excellence Framework, the system for assessing research quality and a great source of stress to university academics, was postponed until further notice. This is explicitly to accommodate the disruption universities face as the result of COVID-19. The announcement was met with widespread relief across the sector.

- Commercial entities in academia, such as journals and publishing houses, are forgoing paywalls and making their entire content on COVID-19 open access, a move that until now was subject to implementing a highly complex arrangement such as the open-access movement Plan S. In many cases, this lowering of

the paywall applies beyond papers related to COVID-19, providing essential reading material to relieve the inevitable boredom and the anxiety of being locked in our houses, proving that if there is a will, there is a way.

• Finally, automated messages from journals chasing peer reviews have been eliminated and extensions offered. This crisis has highlighted how automated computerized actions are inhuman during a time when researchers are constrained by new methods of working and teaching, illustrating the need for general humanity in our research behaviour.

In my view, kindness in research culture is possible beyond COVID-19 – and efforts to make the research world a nicer place to work were happening long before the coronavirus crisis. In January 2018, *Nature* reported on a ‘Kindness in Science’ workshop held at the University of Auckland, New Zealand (see *Nature* 553, 367–369; 2018). And in September last year, the UK charitable research funder Wellcome launched a “reimagine research” campaign. A summit on the topic, due to be held in London on 18 March, was postponed until later in 2020 because of COVID-19.

For change to happen, it’s important to recognize why previous calls for a culture shift were heeded but rarely put into practice. Why was this? A reluctance to change,

**“The pandemic has shown that with the right impetus, change can happen.”**

lack of reward and recognition, personal ambition? Perhaps it goes into the ‘too hard’ basket because of institutionalized practices (human, managerial and technological) that are, by nature, resistant to change.

The increased competitiveness and pressure towards research and university excellence (highlighted during the February 2020 strikes by UK academics) drive a culture in which the mental health of researchers is compromised and discriminatory behaviours are overlooked, and then embedded in research practice.

The COVID-19 pandemic has shown that with the right impetus, change can happen. There are still many elements of research culture that are in need of change, such as gender and racial disparities, bias and the need for more opportunities to foster the careers of young researchers, to name a few. In the meantime, we could use the momentum of COVID-19 to firmly embed kindness into research practice, extending greater goodwill beyond this temporary situation.

**Gemma Derrick** is a senior lecturer in higher education at Lancaster University, UK.

# SEEKING SOLACE IN SCIENCE FUN

Videos, podcasts and light reading can help to lift the mood. **By Atma Ivancevic**

“Find your passion” is, I think, one of the most overused and unhelpful clichés spouted on university campuses. The idea that everyone has a fixed set of interests is harmful, because it encourages us to drop subjects that don’t instantly bedazzle us. It creates the illusion that the perfect project will always be enjoyable. In reality, every scientific venture comes with its own set of challenges. And choosing a PhD or postdoctoral project is influenced as much by factors such as location and availability as by topic.

Although PhD work can spark passion, it has a dark side. *Nature’s* November 2019 survey of more than 6,000 graduate students found that although most respondents were satisfied with their decision to pursue a PhD, more than one-third had sought help for anxiety or depression caused by their PhD studies (see *Nature* 575, 406–406; 2019). The fact that students keep enrolling in PhD programmes, despite the documented experiences of past students, perhaps points towards a love of science and discovery that doesn’t always make its presence felt in the day-to-day work of a PhD student.

Scientists everywhere have faced unprecedented challenges in the past few weeks as laboratories shut down and employees shifted to working from home. For some of us, these changes occurred within the space of a week, meaning that all experiments and cell lines had to be thrown out or frozen down at only a few days’ notice. Having worked from home for two weeks now, I’ve realized that life as a remote scientist is full of Zoom lab meetings, Zoom hacking sessions, Zoom catch-ups and occasional Zoom happy hours (complete with ‘quarantinis’). As a computational biologist with no children or teaching obligations, the transition to working from home should theoretically have been easy. As a human with increasing worries and a limited attention span, the current situation is, at best, distracting.

Personally, I’ve stopped counting how many ‘productive’ hours I have in a day. Instead, I’ve found it helpful to stay up to date with aspects of science that interest me, even if my current topic or task is not going smoothly. I find the reminder that I’m still ‘into’ science, which I often get from outside

my day-to-day work, a reassurance during these uncertain times. Here are some of the resources I use whenever I need a science pick-me-up.

## YouTube

Learning scientific concepts through video can be entertaining, informative and sometimes hilarious. Here are some of my favourite channels.

Simone Giertz ([go.nature.com/2kr4vbn](https://go.nature.com/2kr4vbn)). Formerly known as the Queen of Shitty Robots, Giertz has made videos that showcase the delight of making useless things, as well as not-so-useless things such as a Tesla turned pick-up truck – Truckla.

A capella science ([go.nature.com/2xve6bf](https://go.nature.com/2xve6bf)). Combining deep science with sweet harmony, Tim Blais creates musical summaries of a huge range of science topics including, but not limited to, evolutionary developmental biology, quarks and exoplanets.

True Facts by zefrank1 ([go.nature.com/3awzxzm](https://go.nature.com/3awzxzm)). As an Australian in the United States studying mammals, I often get asked about kangaroos and koalas. I always recommend his marsupials explanation ([go.nature.com/2vuyun](https://go.nature.com/2vuyun)).

## Podcasts

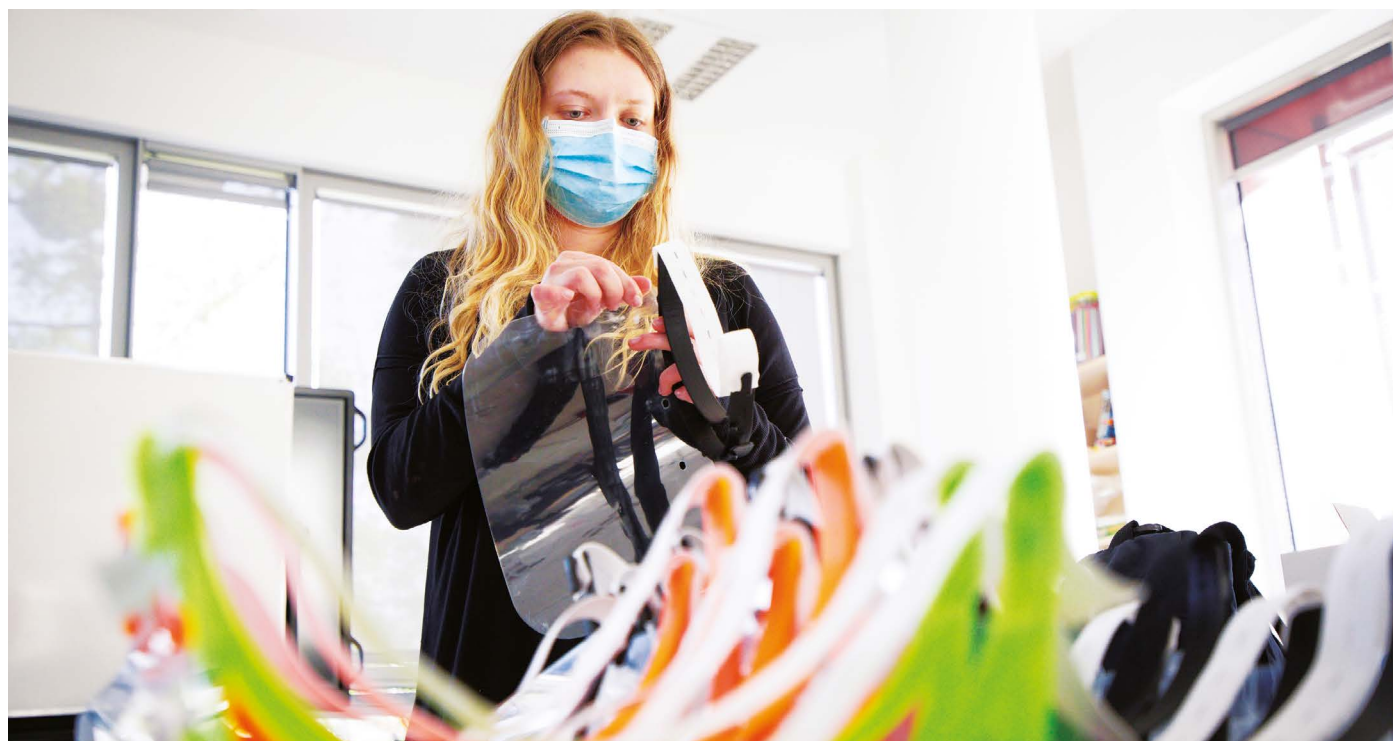
Currently, my go-to science podcasts are Buffs Talk Science and Science Vs.

## Layperson-friendly reading

I often read summaries (across all fields, not just genetics) by Ed Yong (*The Atlantic*) and Carl Zimmer (*The New York Times*) before pursuing the underlying publications.

Science is meant to be fun. Taking a break to program with pictures or revel in an interesting discovery about barnacle sex can make the self-isolation of working from home more bearable. Whether that means experiencing Eminem-inspired physics a capella or discovering a new podcast, we could all use some light-hearted science right now.

**Atma Ivancevic** is a bioinformatics postdoc at the University of Colorado Boulder. She studies ancient retroviruses: genetic parasites that invaded our ancestors millions of years ago and still have a role in modern diseases.



A student at the University of Warsaw assembles 3D-printed protective masks.

# OPEN SCIENCE TAKES ON COVID-19

Data sharing and hobbyists are being harnessed to combat the pandemic. **By Mark Zastrow**

**W**hen reports emerged in late 2019 of an outbreak of a new coronavirus centred in Wuhan, China, researchers at the virological-analysis website Nextstrain were ready. The open-source project tracks the spread of viruses through genetic variations in the sequences that scientists find. After five years of development and operation, Nextstrain had team members on three continents who could continuously refresh the analysis, 24 hours a day.

What they didn't know was whether researchers would share their data. "You just never know what level of detail is going to be allowed to come out," says Emma Hodcroft, a Nextstrain developer and molecular epidemiologist at the University of Basel in Switzerland.

But since 11 January, when a team led by Zhang Yong-Zhen at the Shanghai Public Health Clinical Center, China, shared the first genome sequence of the SARS-CoV-2 virus, the volume of data has skyrocketed. By the end of March, Nextstrain was receiving anywhere

from 50 to 200 sequences a day from laboratories around the world, and was running its analysis of virus evolution every few hours. "The volume that we're getting right now, this is totally unprecedented," says Hodcroft.

Nextstrain is just one example of how an open ethos has driven the scientific response to the COVID-19 pandemic. Academics, online data repositories and home hobbyists with 3D printers are adopting new practices of rapid data sharing and collaboration that are appropriate to the urgency of the crisis. Many hope it will change the way science is done even after the pandemic subsides.

## Do it yourself

Perhaps nowhere is that open ethos clearer than in the way do-it-yourself (DIY) and 'maker' communities have stepped up. As soon as it became clear that health systems around the world were at risk of running out of crucial equipment to treat people with COVID-19 and protect medical workers, DIY-ers set about trying to close the gap.

Facebook groups such as Open Source

COVID19 Medical Supplies, which has more than 70,000 members, have become dispatch centres, through which hospital workers seek volunteers to print or make supplies, and volunteers trade tips on what materials to use and where to source them, and on sterilization procedures.

The coronavirus crisis plays to 3D printing's strong points – rapid prototyping and the ability to produce parts on demand anywhere in the world. Prusa Research, a manufacturer of 3D printers in Prague, has designed a frame for a face shield that is meant to be worn outside a mask or respirator to protect against infectious droplets. The company says it has the capacity to produce 800 shields per day, and tens of thousands of the devices are already protecting health-care workers in the Czech Republic. But because the company made its designs open-source, they are also being made around the world in maker spaces and homes.

Formlabs, a 3D-printer manufacturer based in Somerville, Massachusetts, leads another project that has reached production: printing nasal swabs for COVID-19 test kits. Unlike common cotton swabs, nasal swabs must have a rod that is long and flexible enough to reach deep into the nose, to the upper throat. The swabs were designed by doctors at the University of South Florida in Tampa and the Northwell Health hospital system in New York, using printers purchased from the company to produce test versions. "They are prototyping it themselves, which is crazy and really awesome," says Formlabs's chief product officer, Dávid Lakatos. And whereas conventional swabs feature a bushy tip coating of



nylon flock, the doctors devised a tip with an intricately textured pattern that is 3D-printed.

But unlike face shields, these parts are beyond the capabilities of most printers used by hobbyists. “If someone tried to print the swabs on a hobbyist printer, they can really do harm” in a clinical setting, says Lakatos.

Under US regulations, commercial manufacturing of nasal swabs must take place in a facility that has been registered with the US Food and Drug Administration (FDA). Formlabs has its own registered lab in Millbury, Ohio, with 250 printers (each costing about US\$3,500) that can print 100,000 swabs a day.

### The right tool

Other 3D-printed and DIY projects seek to provide everything from protective face masks for medical workers to door handles that can be opened using an elbow – helping health-care staff to avoid contaminating their hands – and ventilators for people who are critically ill. Among the furthest along in development are the OpenLung ventilator – a collaboration between groups based in Toronto, Canada, and Dublin – and the MIT Emergency Ventilator developed at the Massachusetts Institute of Technology in Cambridge. But manufacturing of such devices is still subject to regulatory approval. The MIT team told *Nature* that “approval would be sought by a manufacturer that ultimately adapts and makes a device inspired by the open-source reference material”. On 17 April, the first such device, called Spiro Wave, received Emergency Use Authorization from the FDA, making it available for use during the crisis; New York City has already ordered 3,000 units.

In the United Kingdom, a collaboration between University College London (UCL), the UCL Hospital and the Mercedes Formula One racing team has reverse-engineered and optimized a ‘continuous positive airway pressure’ device. The design has been approved by UK authorities for use during the COVID-19 pandemic and made available at no cost to manufacturers and researchers. The National Health Service has ordered up to 10,000 units. And in Nigeria, Yunusa Mohammed Garba, a researcher at Gombe State University, has built a positive-pressure ventilator from hobbyist and second-hand components, for use in the northeastern Gombe state, a resource-constrained part of the country. Nigeria has a population of about 200 million, yet it might have fewer than 500 ventilators. Garba’s design is currently being optimized and tested for use at the Federal Teaching Hospital Gombe, which plans to obtain two devices. “At the moment [the ventilator] can only be used in the ambulance,” Garba says. “We are currently using funding from the government to build an upgraded version of the ventilator that can be fully utilized in the hospitals.”

Still, even proponents of 3D printing find

some of the projects potentially dangerous. “It’s both inspiring and extremely scary,” says Lakatos. Formlabs, for instance, investigated face-mask designs and produced numerous prototypes before recommending against 3D printing them. “The [3D-printed] face masks that I’m seeing, those designs are absolutely not sealing anything,” says Lakatos. “And I think they may be even giving a false kind of confidence to people.”

Following discussions with clinicians, Formlabs has instead been recommending a DIY respirator design produced by Boston Children’s Hospital in Massachusetts that repurposes off-the-shelf parts, including ventilator filters and a face mask used for administering anaesthetic gas. “It seems to be a much better solution than trying to do it with 3D printing,” says Lakatos.

One of the most widespread open-source face-shield designs eschews 3D printing entirely. The project began in March with the University of Wisconsin Makerspace in Madison, which worked with Midwest Prototyping, a 3D-printing company in nearby

### “It’s both inspiring and extremely scary.”

Blue Mounds, to produce them. But after bringing in Jesse Darley at the Madison office of Delve, an engineering design firm, the group decided to change tack. Instead of 3D printing, the frames and straps of the resulting ‘Badger Shields’ (named after the university’s mascot) are made from elastic and foam that can be purchased off-the-shelf in bulk form, and cut down either by machine or by hand. Darley says such components can be made in 20 seconds, compared with several hours through 3D printing.

The Madison group has already received orders for five million shields. To meet that demand, manufacturers have stepped in to help, including Ford Motor Company, which Darley says has tweaked the design for mass production and can make around one million shields per week.

### Where credit’s due

The open ethos is influencing other aspects of the pandemic response, too. More than 2,000 articles on COVID-19 have been posted in the preprint archives bioRxiv and medRxiv, according to biomed-sanity.com, a site that aggregates preprints related to the pandemic. Numerous COVID-19 data sets are available on the code-sharing site GitHub, including the data underlying Johns Hopkins University’s widely used COVID-19 case-tracking dashboard. So, too, are reviews of the COVID-19 literature by researchers at the Icahn School of Medicine at Mount Sinai in New York City, and separately by a collaborative project led by computational biologists Halie Rando and

Casey Greene at the University of Pennsylvania in Philadelphia.

Yet it wasn’t a given that researchers would embrace openness early in the outbreak: data that are made public can be difficult to publish through conventional channels later. And multiple news reports have suggested that health workers and researchers in China were initially subjected to government limits on what information they could release. But when Chinese researchers uploaded the first genome sequence of the SARS-CoV-2 virus to the online repositories virological.org and GenBank, they opened the floodgates for more sequences from China and from the rest of the world, Hodcroft says. “I am very grateful for the scientists who took this risk, because I think this set the precedent for the rest of the epidemic.” Given that the outbreak was initially confined to China, had those researchers not done so, “we might have completely different pictures that might be incorrect” she says.

Hodcroft hopes that these collaborative practices will carry over to research on other viruses and seasonal outbreaks. Not all labs have the equipment and personnel to sequence a viral genome, and even for those that do, the work requires time and money. But if more take that step – even if only every couple of weeks – she says it should be possible to track outbreaks in greater detail, using mutations as markers to better understand their geographical spread.

The pandemic could also bring lasting changes in how medical equipment is developed, produced and distributed. Lakatos would like to see hospitals have their own 3D printers as an emergency back-up to produce crucial equipment such as nasal swabs. And publicly available designs for parts such as face shields could make it easier to overcome breakdowns in international supply chains, allowing for more flexible, distributed manufacturing, says Darley.

Soon after releasing the Badger Shield plans, Darley was contacted by a company in Bonner, Montana, he had worked with that makes cycle rickshaws. The firm, called Coaster Cycles, had laid off or drastically cut hours for nearly all its workers because of the pandemic, but after seeing the open-source plans, it won a contract to supply shields to health systems spanning six US states – and hired back its workforce to produce them. Chief executive Ben Morris says the company eventually hopes to sell one million face shields. “That’s the power of open source,” says Darley. “It allows a family to make a few [units], or a manufacturer to make thousands, or hundreds of thousands.”

**Mark Zastrow** is a writer based in Seoul, South Korea.

Additional reporting by Abdullahi Tsanni in Abuja, Nigeria.



## Where I work Israel HersHKovitz

I interview people who lived in the past. I'm an anthropologist at Tel Aviv University in Israel, but I call myself a biohistorian. I'm trying to understand human history, not from human artefacts, temples or big walls surrounding old cities, but from bones. They record every important aspect of your life: what you eat, your ailments, your broken hand at age five.

My main interest is the origin of our own species, *Homo sapiens*: how we evolved, and how we interacted with other *Homo* groups, such as *Homo neanderthalensis*. In this photograph, taken in the preservation laboratory where we first clean our finds, I'm using a pneumatic drill to slowly expose the bones of a human who lived between 100,000 and 120,000 years ago. I and a colleague discovered them in a small cave called TinsHemet, located near the city of Shoham in central Israel. It is also known as Mugharet Al Watwat in Arabic, meaning 'cave of the bat'. The block, which my team removed from the cave, is a half-tonne piece of breccia, an accumulation of sedimentary fragments of rock or mineral debris.

This is probably one of the most important human fossils ever found – the earliest intentional burial that we know of. The skeleton is fully articulated and lies on its side in the fetal position, which suggests that the burial was organized. We don't know yet if it's a modern human, an archaic *Homo* or a Neanderthal, because it will take at least 18 months to expose the bones.

Right now my lab is closed; nevertheless, I continue to work from home at full speed. Anthropology requires the collaboration of scientists from different disciplines, and we do Zoom meetings almost every day.

In this work, your fossil becomes your best friend. You want to know them, to get excited about every discovery and new morphological trait. It's very relaxing to work slowly in a quiet environment, not disturbed by people or external noise. This is what I love the most.

**Israel HersHKovitz** is director of the Dan David Center for Human Evolution and Biohistory Research at Tel Aviv University.

**Interview by Josie Glausiusz.**

Photographed for *Nature* by  
Corinna Kern.